



Università degli Studi di Napoli *Federico II*

DOTTORATO DI RICERCA IN FISICA

Ciclo XXIX

Coordinatore: prof. Salvatore Capozziello

Characterization of Active Galactic Nuclei through Photometric and Spectroscopic Variability

Settore Scientifico Disciplinare FIS/05

Dottorando

[Demetra De Cicco](#)

Tutori

Prof. [Maurizio Paolillo](#)

Prof. [Giuseppe Longo](#)

Prof. [William N. Brandt](#)

Anni 2014/2017

Contents

List of Figures	iv
List of Tables	vii
Introduction	viii
1 Overview of Active Galactic Nuclei	1
1.1 AGN Taxonomy	4
1.2 A Unified Model for AGNs	8
1.3 AGN Variability	13
1.4 AGN Selection Techniques	17
1.5 Broad Absorption Lines in QSO Spectra	20
 Variability-selected Active Galactic Nuclei in the VST Surveys of the COSMOS Field and the Chandra Deep Field South	 28
2 The VST Multi-epoch Survey of the COSMOS Field	29
2.1 The VST and the SUDARE Survey	30
2.2 The SUDARE-COSMOS Dataset	32
2.2.1 The <i>VST-Tube</i> Pipeline	34

3	Selection of AGN Candidates	38
3.1	Catalog Extraction	39
3.2	Aperture Selection and Correction	41
3.3	Defects and Aesthetic Artifacts	44
3.4	The Sample of AGN Candidates	46
4	Validation and Characterization of the Sample of AGN Candidates	54
4.1	X-ray Counterparts	55
4.2	Spectroscopic and Photometric Classification	60
4.3	Color-based Classification	63
4.3.1	Optical-IR Diagnostic	63
4.3.2	Mid Infrared Diagnostic	65
4.4	The Structure Function of Confirmed AGNs	67
5	Extension of the Variability Analysis to the CDFS	77
5.1	Main Results	79
6	Discussion	84
 Broad Absorption Line Quasi-Stellar Objects: Disappearance of the C IV Line Trough in the SDSS-BOSS Sample		93
7	Sample Selection and Data Reduction	94
7.1	The Baryon Oscillation Spectroscopic Survey	94
7.2	Selection of QSOs and Spectra	95
7.3	Spectral Data Reduction	98
7.3.1	Extinction Correction	99

7.4	Continuum Fit	99
8	Statistical Analysis of Disappearing BAL Troughs	103
8.1	Identification of Disappearing BAL Troughs	103
8.2	Statistical Properties of Disappearing BALs	107
8.2.1	Velocity Distributions	109
8.2.2	Equivalent Widths and Coordination in BAL Variability	115
8.3	Comparison to Results from Filiz Ak et al. (2012)	119
9	Discussion	127
	Conclusions	131
	Bibliography	134

List of Figures

1.1	M-sigma relation	3
1.2	Gallery of AGN optical spectra	9
1.3	Schematic view of the AGN unified model	11
1.4	AGN spectral energy distribution	14
1.5	QSO structure function	17
1.6	BPT diagram	20
1.7	Schematic view of QSO wind structure	23
1.8	BAL QSO spectrum	25
2.1	VST image of the COSMOS field	37
3.1	Source with irregular shape in a defected epoch	42
3.2	Flux growth curve as a function of aperture for a reference star	44
3.3	Example of region to be masked	45
3.4	Magnitude distribution of sources in the epoch with the best seeing	47
3.5	Light curve rms as a function of the average magnitude	49
3.6	Examples of variable AGN candidates assigned to different quality classes, with their light curves	52
3.7	Location of the AGN candidates in the VST-COSMOS field .	53

4.1	The COSMOS field as imaged by the VST, <i>Chandra</i> , and XMM observatories	56
4.2	Hard X-ray luminosity vs. redshift for the X-ray AGN candidates	59
4.3	Optical magnitude vs. X-ray flux diagram	61
4.4	r - z vs. z - k diagram	72
4.5	IRAC colors of galaxies with different AGN contributions . . .	73
4.6	MIR diagram (COSMOS)	74
4.7	Structure function of confirmed AGNs	75
4.8	Structure function as a function of black hole mass and Eddington ratio	76
5.1	The CDFS as imaged by several observatories	78
5.2	Optical/IR diagnostic separating stars from galaxies	82
5.3	MIR diagram (CDFS)	83
6.1	Light curve rms of X-ray sources as a function of the average magnitude, for different observing baselines	91
6.2	Fraction of confirmed AGNs as a function of the observing baseline	92
7.1	Examples of continuum fits	102
8.1	Examples of C iv BAL disappearance	111
8.1	Examples of C iv BAL disappearance (cont.)	112
8.2	Redshift distribution for the P_4 sample and the main sample .	113
8.3	Distribution of the rest-frame time difference between epoch pairs where disappearance is observed	113
8.4	Velocity distributions for the P_4 sample	114

8.5	EW distributions	117
8.6	EWs at two different epochs for the ND sample	118
8.7	Fractional EW variation as a function of velocity offset for the ND sample	120
8.8	BAL QSOs identified by Filiz Ak et al. (2012) and not recov- ered from our selection	123
8.8	BAL QSOs identified by Filiz Ak et al. (2012) and not recov- ered from our selection (cont.)	124
8.9	Velocity distributions for the P_8 sample	125
8.10	Velocity distributions for the FA sample	126

List of Tables

2.1	COSMOS dataset	36
6.1	Confirmed sources in the secure sample	90
8.1	Numerical information about the main sample and the P_4 and P_8 samples	106
9.1	Comparison with the main numerical findings by Filiz Ak et al. (2012)	129

Introduction

Nowadays we generally accept the idea that the emission from active galactic nuclei (AGNs) is powered by an accreting supermassive black hole (SMBH) at the center of a galaxy. It is also well known that most –probably all– galaxies go through an active phase at some point in their lives, and that several empirical relations connect black hole and host galaxy properties: this suggests that a tight feedback between the evolution of the black hole and the host galaxy exists; hence, in order to go deeper into galaxy evolution, it is crucial for us to learn more about the formation and evolution of the black holes residing in their centers.

Variability is a defining property of AGN emission at all wavebands, and concerns both continuum and broad-line emission. It is generally attributed to instabilities in the AGN accretion disk, together with changes in the accretion rate. Since the extent of variations in different wavelength ranges is not the same, variability measurements can help understand the underlying emission mechanism, constraining the size and structure of the emitting region.

The present work investigates AGN variability from two different perspectives. The first part of the project tests the efficiency of optical variability as a tool to select AGNs, since optical continuum variability seems to be a universal feature of broad-line AGNs on timescales from months to years, with

variations generally ranging from 1% to 10% of the magnitude, but also much larger (50% of the magnitude) in some cases. Testing techniques for AGN identification based on data from ground-based telescopes is of great relevance in the framework of current and future wide-field surveys (e.g., Dark Energy Survey, Large Synoptic Survey Telescope), since we will need reliable methods to detect and classify the wealth of sources they will provide.

The second part of the project investigates the variability of broad absorption lines (BALs) in quasi-stellar object (QSO) spectra, in order to deeply understand the physics and structure of AGNs. BALs originate from out-flowing winds along our line of sight; winds are thought to originate from the accretion disk, in the very proximity of the central SMBH; we generally think that they are responsible for a triggering of the accretion mechanism onto the SMBH, as they remove angular momentum from the disk and, since they evacuate gas from the host galaxy, they also play a leading role into galaxy evolution. Several works show that BAL equivalent widths can change on typical timescales from months to years. Such variability is generally attributed to changes in the covering factor (due to rotation and/or changes in the wind structure) and/or in the ionization level. We investigate BAL variability, focusing on BAL disappearance, in a sample of more than 1500 QSOs –the largest sample ever used for such an analysis– to gain insight into the structure and co-evolution of the SMBH and the host galaxy.

Chapter 1

Overview of Active Galactic Nuclei

Historically, a galaxy is defined “active”¹ if its nuclear emission can be attributed to physical processes of non-stellar origin. In such a case, we say that the galaxy is hosting an active galactic nucleus (AGN). If an AGN is bright enough to be resolved, and not obscured by the presence of dust, its typical bolometric luminosity L_{bol} ranges $10^{42} - 10^{48}$ erg s⁻¹, and thus can be up to 10^4 times higher than the bolometric luminosity of a typical non-active galaxy. The active region inside a galaxy is generally very compact, having linear size on the order of 1 pc, while its angular size, beside the intrinsic properties, depends on the observing wavelength as well as on redshift and on the luminosity contrast of the nucleus with respect to the rest of the galaxy (e.g., Krolik, 1999).

Active galaxies are characterized by broadband spectra exhibiting local maxima, and emission in the X-rays and in the radio is several order of

¹Star-forming galaxies are considered active as well; in the case of AGNs, the active part is the central black hole (as discussed in the following), rather than the galaxy itself.

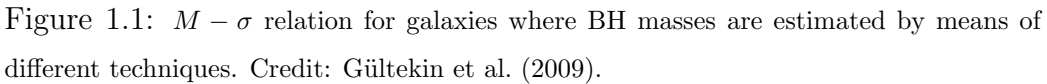
magnitudes larger than in non-active galaxies. The continuum emission can be roughly described by a power-law:

$$F_\nu \propto \nu^{-\alpha} , \quad (1.1)$$

F_ν being the spectral flux density, while α is the spectral index, which generally ranges 0 – 1 between X-ray and optical wavelengths. In normal galaxy spectra, emission is essentially ascribable to the presence of stars, and hence is typically black-body emission, mirroring different stellar temperatures; atoms and molecules in stellar atmospheres produce absorption features while, when star-forming H II regions are present, emission features are produced as the result of the heating of gas which then re-emits radiation at specific wavelengths. AGN emission is rather the combination of a continuum, resulting from different physical processes (Compton scattering, multi-temperature black body emission, synchrotron emission), and emission features which are generally broader and much more prominent than the ones commonly observed in the spectra of non-active galaxies. Typical AGN optical/ultraviolet (UV) emission lines include $\text{Ly}\alpha$, lines from the Balmer series, the C IV 1549 Å doublet, the Mg III line at 2799 Å, the [O III] forbidden line at 5007 Å, and iron line blends.

Nowadays it is widely accepted that the engine powering an AGN is an accreting supermassive black hole (SMBH) in the center of the galaxy, and very likely all galaxies host a SMBH in their center and undergo an active phase during their lives; the typical mass of the central SMBH is $M \gtrsim 10^7 - 10^8 M_\odot$ ². A number of empirical laws relating properties and physical quantities of the central SMBH and the host galaxy have been discovered through the last decades: the most remarkable example is probably the connection between the mass M of the nuclear black hole (BH) and the velocity dispersion σ of

² M_\odot is the mass of the Sun and corresponds to $\approx 2 \times 10^{33}$ g.



Other empirical relations connect the BH mass to the galaxy luminosity (e.g., Kormendy & Ho, 2013; Kormendy & Richstone, 1995) or to the galaxy bulge mass (e.g., Magorrian et al., 1998); moreover, evidence exists for a

co-evolution, through cosmic time, of star formation rate (SFR) and the quasi-stellar object (QSO; see further) luminosity function (e.g., Fiore et al., 2009). All this supports the mentioned hypothesis about the existence of a feedback between SMBH and galaxy evolution and, considering that most massive galaxies host a SMBH, an extensive knowledge of the black hole demography is an essential premise to gain insight into galaxy evolution as a whole.

1.1 AGN Taxonomy

Today we label as AGNs a variety of heterogeneous sources which, historically, were given different names: this mainly happened because observations were obtained in different wavelength ranges, thus highlighting some features rather than others, although intrinsic differences played a role as well. As a consequence, the attribution of an AGN to a specific subclass is not to be intended as rigorous, and different subclasses can partially overlap. Moreover, the classification of a source can change in time due to variability (see Section 1.3).

A primary classification of AGNs emerges from the comparison of the active nucleus luminosity L_{nuc} to the host galaxy luminosity L_{gal} , both measured in the ultraviolet/optical/infrared (IR), and introduces three subclasses of objects:

- QSOs, for which $L_{\text{nuc}} \gtrsim L_{\text{gal}}$;
- strong AGNs, for which $L_{\text{nuc}} \lesssim L_{\text{gal}}$;
- weak AGNs, for which $L_{\text{nuc}} \ll L_{\text{gal}}$.

We could state that QSO typical bolometric luminosities are on the order of 10^{46} erg s $^{-1}$, while strong AGNs usually have $L_{\text{bol}} \approx 10^{44}$ erg s $^{-1}$, and weak AGNs are characterized by bolometric luminosities around 10^{42} erg s $^{-1}$. Differences in luminosity mainly reflect differences in the accretion rates (but also in the masses) characterizing the various AGNs.

An alternative classification is based on the relevance of radio emission in the source spectral energy distribution (SED): traditionally, AGNs are classified as radio-loud when they emit a significant amount of energy in the radio band, and radio-quiet otherwise³. Different classifications of radio-loud AGNs exist, based on the spectral index of the source power-law spectrum (see Equation 1.1), or on morphology (e.g., Fanaroff-Riley classification, where the distance between the two brightest spots in the radio image of a source plays a key role).

In what follows we are going to sketch the defining features of the main traditional AGN subclasses (e.g., Peterson, 1997; Krolik, 1999); we point out that a number of additional classes are nowadays generally included in the AGN family.

Seyfert galaxies⁴ are mostly spirals characterized by a high surface brightness nucleus (B band magnitude $M_B \geq -21 + 5 \log h_0$ mag, where $h_0 = H_0/100$ km s $^{-1}$ Mpc $^{-1}$, H_0 being Hubble's constant); the host galaxy is generally resolved, and spectra typically exhibit prominent emission lines, including strong high ionization lines and coronal lines (e.g., [Fe X], [Fe XI], [Si IX], [Si X]). Depending on the width of the emission lines observed in their optical spectra, Seyfert galaxies are split into Type 1 and Type 2: in both

³Nowadays the use of the alternative dichotomy jetted/non-jetted is gaining in importance (e.g., Sbarrato et al., 2014, and references therein).

⁴Seyfert galaxies are named after their discoverer C. K. Seyfert who, in 1943, identified six galaxies characterized by a very bright and compact nucleus and analyzed their spectra.

cases we observe narrow forbidden lines (e.g., [O III] 5007 Å) produced in a low density medium (typical electron density $n_e \approx 10^3 - 10^6 \text{ cm}^{-3}$) and having a typical full width at half maximum (FWHM) that corresponds to velocities in the range $10^2 - 10^3 \text{ km s}^{-1}$; the difference between the two classes lies in the presence of broad permitted lines (e.g., Balmer or H II lines) in the spectra of Type 1 galaxies, in addition to narrow lines. Broad lines are characterized by a FWHM generally corresponding to higher velocities (up to 10^4 km s^{-1}) and originate in a denser medium ($n_e \gtrsim 10^9 \text{ cm}^{-3}$) than narrow lines. An additional feature of Seyfert 1 galaxies is a strong blue/optical non-stellar continuum, which is absent, or weak and hard to disentangle from the stellar continuum, in the spectra of Seyfert 2 galaxies. Weak absorption lines ascribable to giant stars in the host galaxy may also be detected in the spectra of both types of Seyfert galaxies. The classification as Type 1 or 2 on the basis of the characteristics of the emission lines observed in a spectrum is generally extended to all AGNs; alternative labels for the two classes of sources are broad-line AGNs (BLAGNs) and narrow-line AGNs (NLAGNs), respectively.

Quasi-stellar objects are the brightest AGNs ($M_B \leq -21 + 5 \log h_0$ mag); their nucleus outshines the host galaxy, hence their appearance is stellar-like when observed through an optical telescope, and this explains their name. Their spectra are similar to Seyfert galaxy spectra, but narrow lines, as well as absorption features, are typically weaker (but see Section 1.5, where QSOs characterized by broad absorption lines are introduced). Approximately 5 – 10% of QSOs are radio-loud, and are commonly known as quasars⁵.

⁵After the pronunciation of the acronym QSRS, which stands for “quasi-stellar radio source”. Nowadays the word is mostly used to label QSOs as a whole, regardless the relevance of their radio emission; nonetheless, throughout the present work we prefer to

Blazars⁶ are radio-loud AGNs that we observe face-on (see Section 1.2), characterized by a strong relativistically beamed jet along the line of sight. Blazars are split into flat-spectrum radio-loud quasars (FSRQs; see, e.g., Fos-
sati et al. 1998) and BL Lac⁷ objects, depending on their SEDs and spectral index in the radio band. Both FSRQs and BL Lac objects are characterized by a strong non-thermal broad band continuum; the main difference is that both broad and narrow lines can be observed in FSRQ spectra, while BL Lac objects are characterized by a featureless continuum as the strength of the jet completely dims any spectral line.

Low Ionization Nuclear Emission Regions (LINERs) are faint optical emission-line AGNs characterized by stronger low-ionization lines (e.g., [O I] 6300 Å, [N II] 6548 Å, 6583 Å) than the ones typically found in Seyfert galaxy spectra; according to the original definition, an AGN is a LINER if the oxygen forbidden lines in its spectrum fulfill the conditions $[\text{O II}] 3727 \text{ Å} / [\text{O III}] 5007 \text{ Å} > 1$ and $[\text{O I}] 6300 \text{ Å} / [\text{O III}] 5007 \text{ Å} > 0.32$ (Heckman, 1980). LINERs are the sources showing the weakest form of nuclear activity among AGNs and, according to some works from the literature (e.g., Stasinska et al., 2008; Capetti & Baldi, 2011), they are not powered by an AGN.

Other classes of AGNs include:

- narrow-line X-ray galaxies (NLXGs), which could be considered as faint and reddened Seyfert galaxies characterized by high extinction and high X-ray emission;

keep the distinction, and rather use the more generic term “QSO”. The first QSOs were observed in the ’50s and were powerful radio-loud sources.

⁶The name is quite like the pronunciation of BL Lac (see further); in addition, it recalls the word “blaze”, which is an appropriate association.

⁷BL Lac objects are named after their prototype *BL Lacertae*; their stellar appearance led to the misbelief that they were stars, hence the misleading name.

- starburst galaxies, typically bluish as characterized by a high star formation rate; the presence of O and B stars marks their spectra with H II emission features, as typical of star-forming systems, and radio emission, originating from supernova (SN) remnants, is commonly observed; not all starburst galaxies are AGNs;
- Markarian galaxies⁸, typically exhibiting very strong blue/UV emission and non-thermal continuum;
- ultra luminous infrared galaxies (ULIRGs), characterized by strong far infrared (FIR) thermal emission (typical luminosities $L \gtrsim 10^{12} L_{\odot}$ in the wavelength range $8 - 1000 \mu\text{m}$) originating from dust heated by star formation or from an active nucleus obscured by dust; they generally have a ratio of infrared to optical luminosity $L_{IR}/L_{opt} \gtrsim 10$; like starburst galaxies and LINERs, ULIRGs are considered borderline AGNs.

In Figure 1.2 optical spectra for different AGN types are shown, in order to allow an overview of their main features.

1.2 A Unified Model for AGNs

In the '90s a model capable of gathering under a common label the various classes of objects introduced in Section 1.1 was proposed (e.g., Antonucci, 1993; Urry & Padovani, 1995). The so-called unified model explains the observed differences among the various sources as an effect of the different possible orientations of the sources themselves with respect to our line of sight. According to such model, AGNs are cylindrical symmetry objects,

⁸Markarian galaxies are named after B. E. Markarian, who first analyzed them in 1963.

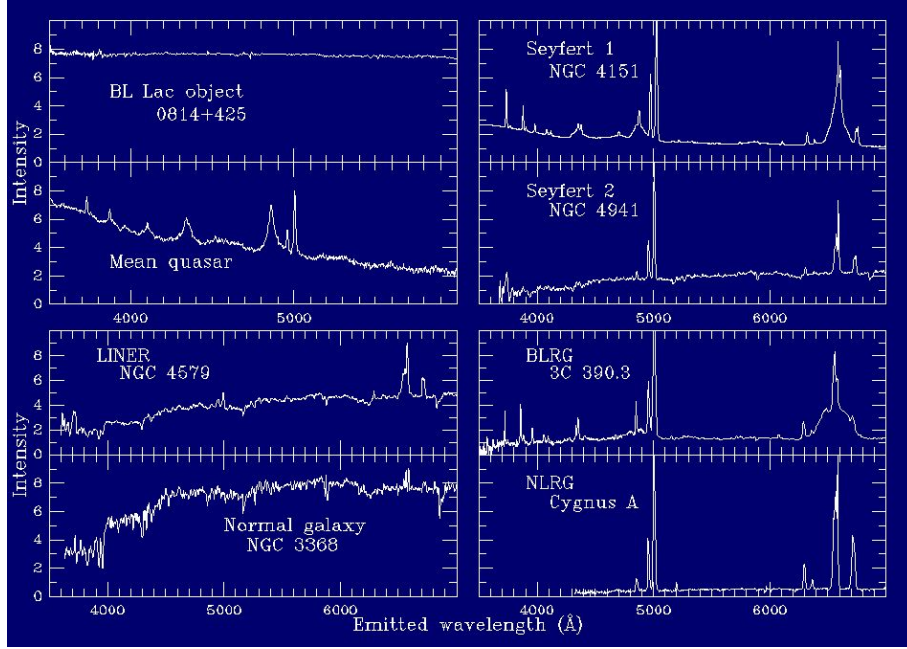


Figure 1.2: Gallery of AGN optical spectra, all shifted to their emitted wavelength scales to ease the comparison. Credit: <http://www.astr.ua.edu/keel/agn/>.

the symmetry axis being the one around which a central supermassive black hole rotates. The SMBH is surrounded by an accretion disk, constituted by matter falling towards the black hole; the matter loses angular momentum through viscous dissipation and turbulent processes, and its gravitational potential energy is converted into radiation, which we detect from X-rays to optical, with decreasing energy as the distance from the central black hole grows. Above the disk a corona forms, which is responsible for the scattering of optical/UV photons from the disk up to X-ray energies (inverse Compton scattering).

In the framework of the unified model, the mentioned UV and optical broad lines characterizing some classes of AGNs originate in the so-called broad line region (BLR), where gas clouds are in rapid motion (velocities up to 10^4 km s^{-1}) in the black hole potential, in the very proximity (≈ 1

pc) of the black hole. Both the black hole and the BLR are surrounded by an optically thick toroidal structure of gas and dust, emitting in the IR; depending on the viewing angle, the torus may obscure the BLR, and this explains the lack of broad lines in the spectra of some AGN types. The torus inner radius is ≈ 10 pc.

At a distance of $\approx 10\text{--}1000$ pc from the center there are gas clouds moving at lower velocities ($10^2 - 10^3$ km s $^{-1}$), constituting the narrow line region (NLR): here the narrower emission lines characterizing NLAGNs originate. The black hole rotation may generate collimated jets of escaping energetic particles relativistically beamed in the direction of the torus axis⁹; the jets are powered by gravitational energy released by the accretion disk and emit synchrotron radiation from radio to γ wavelengths; their light is observed up to ≈ 1 Mpc away from the black hole.

On the basis of the sketched model, the cylindrical symmetry of AGNs, together with their geometry as a whole, accounts for the main differences among the various types of objects: if the AGN is viewed in a direction perpendicular to the torus axis (i.e., edge-on), the torus will obstruct the view of the BLR, and hence in the AGN spectrum there will be few or no broad lines at all, while the broad line component will be more and more evident as the angle between the direction of observation and the torus axis approaches zero (object observed face-on). Therefore, according to the unified model, the various AGN types are nothing other than the same type of object viewed from different angles. Figure 1.3 is a schematic view of the described unified model, in the case of a one-sided jet AGN, while the two panels in Figure 1.4 show where (top) and from which processes (bottom) the various components

⁹Generally there is no particular relation between the orientations of the torus axis and of the host galaxy rotation axis.

of AGN emission originate.

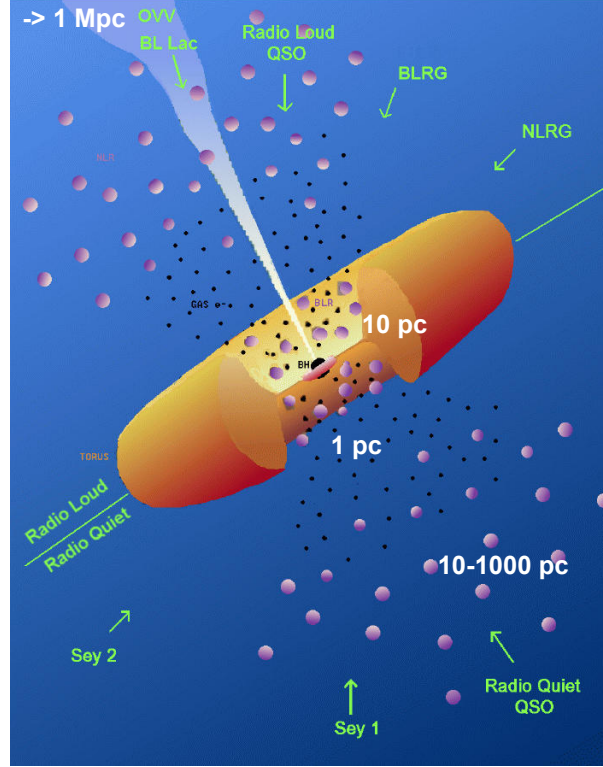


Figure 1.3: AGN unified model (not in scale). Numbers indicate the approximate distance from the central SMBH or the extent of the corresponding region of the AGN. A one-sided jet is shown, in order to get a sense of all the possible observing directions and the corresponding types of observed objects. The various arrows show the directions one is looking through when detecting the corresponding AGN types. Credit: <http://www.cv.nrao.edu/course/ast534/ExtraGalactic.html>, based on the original figure in Urry & Padovani (1995).

A relevant observational evidence which led to the formulation of the unified model concerns narrow lines in Seyfert galaxy spectra: both Type 1 and Type 2 Seyferts exhibit a common set of narrow lines with similar line ratios, and this led to the hypothesis that the same engine powered both types of AGNs. A further step was the discovery of broad lines in the polarized spectrum of NGC 1068: this was surprising as the galaxy is a Seyfert 2, hence

only narrow lines are expected in its spectrum due to the obscuring torus; nevertheless, the light from the BLR is scattered along our line of sight from matter such as free electrons in ionized gas and is highly polarized, so as to be evident in a polarized spectrum although too weak to be detected in a total spectrum (Antonucci & Miller, 1985; Bianchi et al., 2012).

Since the '90s the AGN unified model has been revisited and a number of works discussing revised versions of it can be found in the literature. Here we refer to Netzer (2015), where the most remarkable findings of the last decade are reviewed.

First, it is very likely that most AGN tori are axisymmetric and clumpy systems with a central opening from which two anisotropic cones of ionizing radiation originate; constraining their basic parameters is a tricky task, and our present knowledge essentially relies on empirically derived models –which are far from exhaustive– and on the more encouraging results from hydrodynamic simulations investigating the connection between the disk and the gas inflowing from the host galaxy. The study of the cones allows to constrain the torus boundaries, and relevant results are expected from the use of long-baseline interferometry, as it allows the imaging of AGN inner regions.

Winds (see Section 1.5) play a relevant role in AGN models; yet, no preferred driving mechanism (magnetic forces or radiation pressure) currently exists. The various disk-wind models proposed in the last decades are quite efficient in describing the formation of clumps. There are hypotheses about the BLR being flat in shape, and originating from the disk or from the winds that rise from it. Different covering factors –which are thought to be responsible for anisotropy and absorption line variability (see Section 1.5 and Part II of the present thesis)– are usually taken into account, but their relation with luminosity still needs confirmation.

The AGN family includes a number of classes of sources whose membership is still a matter of debate. Moreover, merging systems seem not to fit into the general unification scheme, as their behavior differs from typical AGN behavior.

1.3 AGN Variability

AGN variability has been known for a long time, even before the nature of AGNs was known. Variability characterizes AGNs in all wavelength bands, affecting both continuum and broad-line emission, with timescales ranging from minutes to years. Variations are aperiodic, and their characteristics depend on a number of factors, such as spectral window, source luminosity, presence of radio jets, and correlations exist among variations in different energy ranges (e.g., Ulrich et al., 1997; Gaskell & Klimek, 2003). In particular, the extent of variations depends on wavelength, and is an indicator of the contribution of the emission in that waveband to the total output from an active galaxy: for instance, optical continuum variability has typical amplitudes of 10^{-2} to 10^{-1} mag (but also much larger, up to 0.5 mag, in some cases; see, e.g., Givon et al. 1999), typical timescales ranging months to years, and it seems to be a universal feature of BLAGNs (e.g., Webb & Malkan, 2000), while magnitude variations around 30% are commonly detected in the X-ray waveband (e.g., Paolillo et al., 2004; Yang et al., 2016; Paolillo et al., in prep.), and they can be even much larger, especially for AGNs with high Eddington ratios.

Variability measurements can contribute in shedding light on the underlying emission mechanism, providing constraints on the size and structure of the emitting region. Knowing the timescale Δt of variability allows to con-

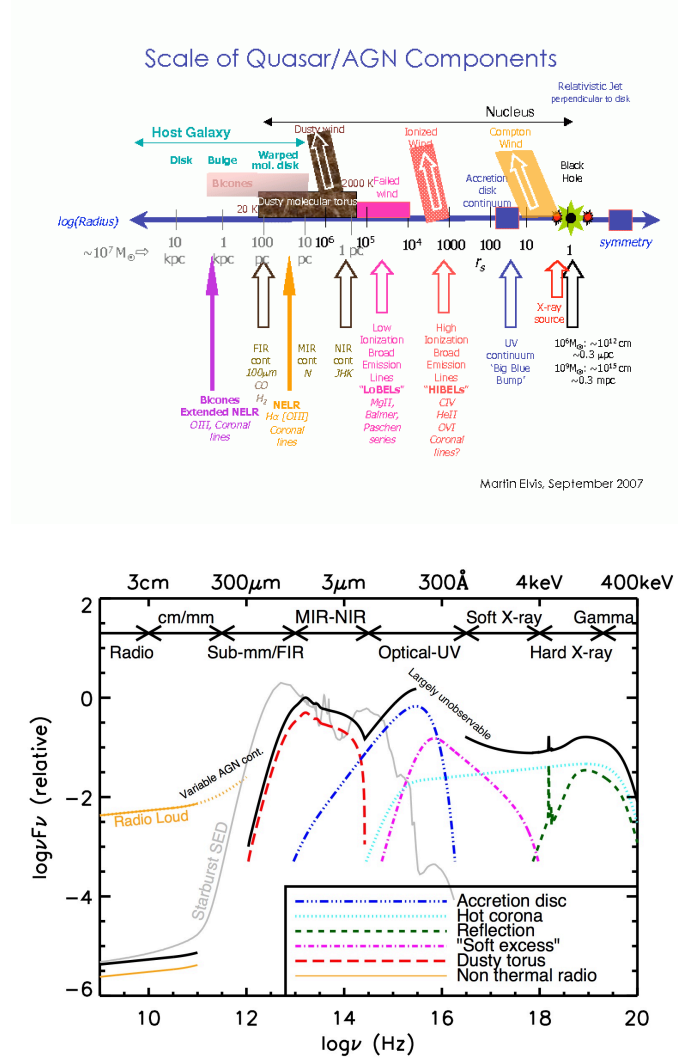


Figure 1.4: *Upper panel*: different components of AGN emission, together with an indication of the distance at which they originate with respect to the central SMBH. Credit: M. Elvis, 2007 (<http://hea-www.harvard.edu/elvis/QuasarScales.pdf>). *Lower panel*: schematic representation of the SED of AGNs, loosely based on the observed SEDs of radio-quiet QSOs (Elvis et al., 1994; Richards et al., 2006). The black solid line represents the total SED, while the colored dashed lines represent the individual components. Credit: C. M. Harrison, PhD Thesis (<http://astro.dur.ac.uk/cpnc25/documents/cmh.thesis.pdf>).

strain the size l (limited to the direction of our line of sight) of the emitting region by means of light travel arguments, which imply that $l \leq c\Delta t$, c being

the speed of light¹⁰.

As an example, a timescale on the order of days ($\Delta t \approx 10^5$ s) corresponds to a size $l \lesssim 10^{-3}$ pc for the region affected by emission variations, while a timescale $\Delta t \approx 1$ h returns $l \approx 10^{-5}$ pc, which approximately corresponds to the Schwarzschild radius for a $10^8 M_\odot$ object.

Most QSOs typically exhibit continuum variations on the order of 10% over months to years, and we already mentioned (Section 1.1) that blazars show even more substantial variations on much shorter timescales.

The leading theories about variability attribute it to instabilities in the AGN accretion disk (e.g., Pereyra et al., 2006) together with other phenomena, such as variations in the accretion rate, presence of obscuring medium, star disruption, and gravitational microlensing (e.g., Aretxaga & Terlevich, 1994).

The analysis of AGN variability at different wavelengths and the study of possible correlations among different spectral windows is nowadays a major field of inquiry.

As for optical variability, which is the subject of the first part of the present work, several attempts to unveil a possible relation to X-ray variability have been made, and there are two leading –and conflicting– theories about it (e.g., Krolik, 1999; Uttley et al., 2006); at present, none of them has been proved right or wrong by the results of the various projects about them. According to one of the theories, optical variability is the result of a reprocessing of X-ray radiation, originating in the AGN corona, which loses energy in the interaction with the surrounding disk and torus. Conversely, according to the other theory, X-ray emission is the one originating from op-

¹⁰We contemplate here an ideal, simplified case where no absorption, nor any other effect, occurs between the AGN and the observer.

tical/UV emission as a result of the Comptonization of optical/UV photons by relativistic electrons. It is likely that both processes occur at the same time in an AGN, as shown by studies of time lags between different bands as a function of temporal frequency.

Optical variability has been extensively used to identify unobscured AGNs in multi-epoch surveys (e.g., Bershadsky et al., 1998; Klesman & Sarajedini, 2007; Trevese et al., 2008; Villforth et al., 2010; Peters et al., 2015, and references therein). The strength of optical variability-based selection lies in the chance to use data provided by ground-based telescopes, which means surveys of large sky areas are available; plus, variability allows to retrieve those AGNs characterized by low X-ray emission and hence not classified as AGNs on the basis of X-ray surveys; moreover, it proves effective in identifying low-luminosity AGNs (LLAGNs) because of the anti-correlation between AGN luminosity and variability amplitude, as discussed in, e.g., Barr & Mushotzky et al. (1986); Lawrence & Papadakis et al. (1993); Cristiani et al. (1996); Young et al. (2012). In the first part of the present work we describe how we used optical variability to identify AGNs in the COSMOS field, which is one of the most extensively and deeply surveyed sky areas (see Section 2.1); our goal is to prove the effectiveness of the method against other traditional approaches.

In Figure 1.5 we show the structure function (SF) of a sample of QSOs in three different optical bands. The SF allows to parametrize the variability of a sample of sources without having to take each of them into account individually; a more detailed description of the SF will be presented in Section 4.4. From the plot it is apparent that variability depends on the specific band, and that larger variations correspond to larger time lags.

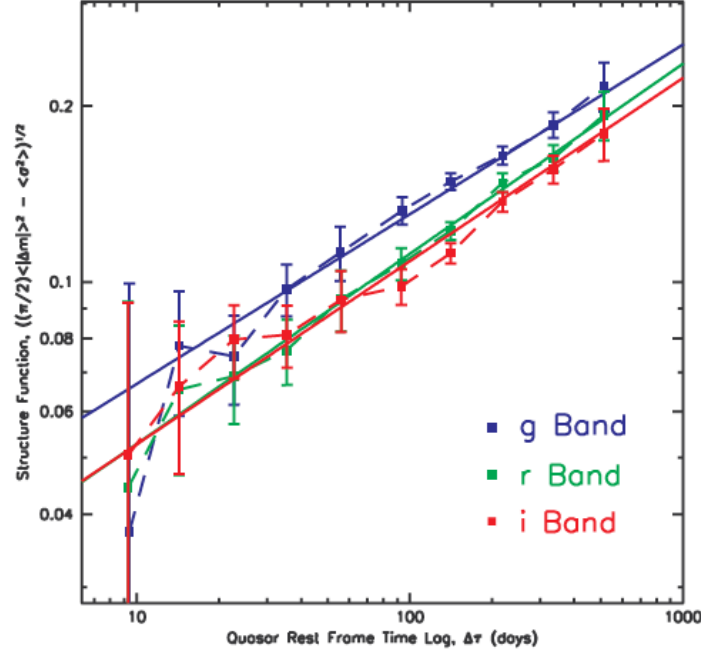


Figure 1.5: Structure function of a sample of QSOs in the g (blue), r (green), and i (red) bands. Power-law fits to each set of data are also shown (solid lines). Credit: Vanden Berk et al. (2004).

1.4 AGN Selection Techniques

Due to the various different features of AGNs, a census of the AGN population taken by means of a single selection technique would not be complete, nor bias-free; as a consequence, a large number of identification methods – including the ones based on variability, introduced in Section 1.3– have been developed and tested in time, each of them aiming at the detection of specific features (e.g., Krolik, 1999; Mushotzky, 2004).

The most reliable and unambiguous signature of an AGN is, undoubtedly, its X-ray emission, whose high penetrating power unearths even AGNs that are obscured at other wavelengths; X-ray emission from AGNs is typically characterized by large-amplitude and high-frequency variability. The

big drawback of relying on X-ray emission to identify AGNs is that space observatories are required, and they are much more expensive and have a more limited field of view (FoV) than ground-based telescopes. We will deal more extensively with AGN identification techniques based on their X-ray features in Section 4.1.

Selection techniques based on colors are widely used when multi-wavelength data in the UV/optical/IR are available (e.g., Fan, 1999). Sources of a different nature have different SEDs, and this reflects into different colors; in particular, AGN emission in the UV (if they are not obscured) and IR bands is remarkable if compared to the emission from stars or normal galaxies in the same bands, hence a diagnostic comparing two or more source colors will allow to identify different classes of objects (e.g., Richards et al., 2001). The main convenience here is that a few images are sufficient to retrieve information about large samples of objects; conversely, the drawback is that the color selection suffers from different biases. First, faint AGNs outshone by their host galaxy will be hidden by typical galaxy colors and hence will not be detected; moreover, the technique suffers from absorption and extinction: the first is due to the presence of dust in the plane of the Galaxy, hence mostly affects low-latitude observations, while the latter can be intrinsic to the AGN itself or be ascribable to the host galaxy, and its weight increases with redshift (e.g., Krolik, 1999). Also, low redshift surveys favor extended sources, as they exclude point-like objects in order to minimize star contamination; on the contrary, high redshift surveys preferentially select point-like sources in order to exclude faint galaxies. In addition, color selection may suffer from a bias due to the non-contemporaneity of observations in different bands. Examples of AGN color selection will be discussed in detail in Section 4.3.

Since some classes of AGNs are characterized by broad and prominent emission features, a number of diagnostics rely on AGN selection on the basis of the properties of the emission lines observed in their spectra, provided that obscuration in the wavelength range of interest is not significant. The most popular example is probably the BPT diagram, named after Baldwin, Phillips, and Terlevich (Baldwin, Phillips, & Terlevich, 1981), where the line ratios $[\text{O III}] \ 5007 \text{ \AA} / \text{H}\beta \ 4861 \text{ \AA}$ and $[\text{N II}] \ 6583 \text{ \AA} / \text{H}\alpha \ 6563 \text{ \AA}$ are compared; line pairs are roughly redshift independent since they are close to each other in wavelength, and they were chosen as they allow to separate AGNs (Seyfert galaxies or QSOs) from star-forming galaxies and from LINERs: AGNs are characterized by high values of both ratios, while star-forming galaxies typically have lower $[\text{N II}] / \text{H}\alpha$ values; LINERs usually have low values of $[\text{O III}] / \text{H}\beta$ if compared to AGNs, and high values of $[\text{N II}] / \text{H}\alpha$ with respect to star-forming galaxies. An example of BPT diagram is shown in Figure 1.6, where different loci defined by different source types can be seen. The drawback in classifying AGNs by means of their spectral features is that spectroscopy is time-consuming, especially when dealing with very large samples or faint sources, and several biases affect the sample completeness: for instance, the set of lines observed in a given wavelength range is redshift-dependent; redshift is also responsible for the restriction by a factor $\frac{1}{1+z}$ of the rest-frame observing window.

A number of other diagnostics making use of colors and/or emission line ratios exist: as an example, the TBT diagram, after Trouille, Barger, and Tremonti (Trouille, Barger, & Tremonti, 2011), combines the rest-frame g - z color and the emission line ratio $[\text{Ne III}] \ 3869 \text{ \AA} / [\text{O II}] \ 3726, 3729 \text{ \AA}$, thus allowing to disentangle AGN-dominated from star formation-dominated galaxies provided that the redshift of the sources is $z < 1.4$ (a constraint defined by

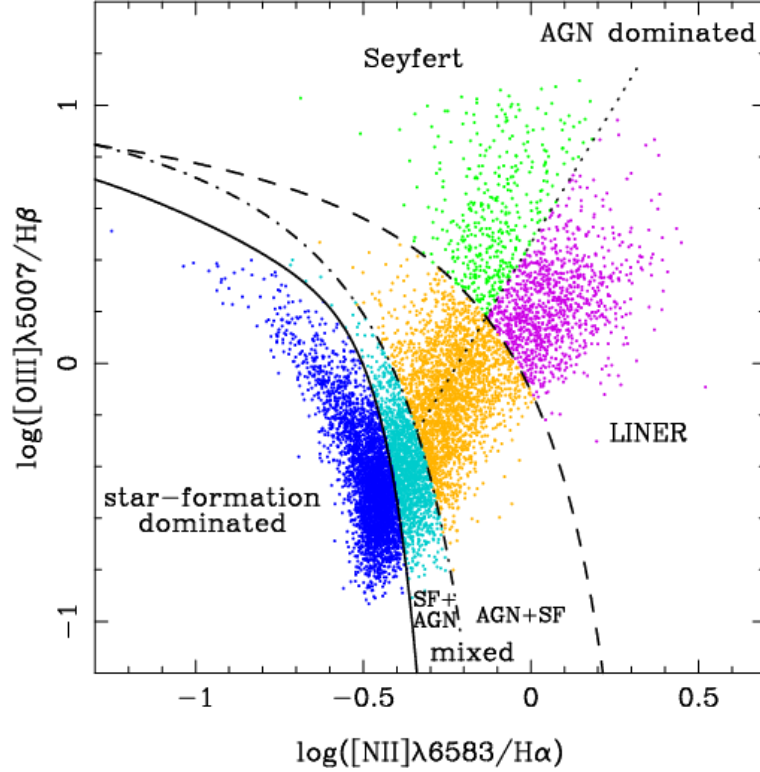


Figure 1.6: Example of BPT diagram where, in addition to the main distinction between star formation-dominated and AGN-dominated galaxies, other regions where different classes of objects place themselves are defined. Credit: Bamford et al. (2008).

the chosen emission lines). For higher redshift sources, near infrared (NIR) spectroscopy is a reliable tool for AGN identification (e.g., Landt et al., 2008).

1.5 Broad Absorption Lines in QSO Spectra

QSO spectra are characterized by prominent emission lines originating from ultraviolet transitions such as C IV, Si IV, N V. In 10% – 20% of optically selected QSOs, to such transitions also correspond intrinsic¹¹ absorption fea-

¹¹The absorption lines at issue are labeled “intrinsic” since they originate from material that is considered to be part of the QSO, and not from the intervening medium between

tures which are blueshifted up to $0.1c$ from the corresponding rest-frame absorption lines and, in $\approx 10\%$ of the cases, additional lower ionization transitions, e.g., Al II, Al III, Mg II, are observed (e.g., Weymann et al., 1991; Murray et al., 1995).

The presence of absorption lines implies the existence of some mechanism capable of transferring a significant amount of momentum from a radiative field to the gas where lines originate. The most widespread theory is that absorption features form because of radiatively accelerated winds originating in the inner regions of the accretion disk, in the immediate surroundings of the central supermassive black hole, at typical distances of $\approx 10^{-2} - 10^{-1}$ pc (e.g., Murray et al., 1995; Proga et al., 2000).

Winds remove from the disk the angular momentum released by the accreting material, thus triggering the accretion mechanism in a substantial way; furthermore, since they evacuate gas from the host galaxy redistributing it in the intergalactic medium, they also affect star formation processes and, as a consequence, galaxy evolution as a whole (e.g., Di Matteo et al., 2005; Capellupo et al., 2012).

Several models have been proposed to describe QSO winds (e.g., Elvis, 2000; Proga et al., 2000) and, according to most of them, the observation of wind absorption lines is a matter of viewing angle, as winds are thought to originate in the equatorial region of a QSO. This constitutes a possible explanation for the lack of detection of the resulting absorption lines in a large fraction of QSO spectra.

Figure 1.7, from Elvis (2000), shows the schematic structure of the wind model there described, derived from observational results: essentially, it proposes a funnel-shaped, biconical structure for BAL outflowing winds. A

the QSO and the observer.

vertical, cylindrical stream originates from some instability in the accretion disk, at a distance on the order of 10^{-2} pc from the central SMBH; due to centrifugal forces combined with radiation pressure, the stream bends outwards radially when the vertical velocity equals the radial velocity, and accelerates to typical BAL velocities. The funnel-shaped structure implies that BALs or narrow absorption lines (NALs) are observed depending on the viewing direction: when a QSO is observed along the flow, BALs are detected (10%-20% of observed QSOs); if the observing direction is towards the vertical part of the funnel, NALs will be visible in the QSO spectrum (40%-45% of the cases); if the funnel is viewed face-on, no absorption lines at all will be observed (the remaining 40%-45% of the cases). The characteristic angles were empirically derived in order to account for the observed fractions of sources exhibiting BALs/NALs/no absorption lines in their spectra. Alternative hypotheses about QSO absorption lines see them as the signature of a peculiar stage of QSO evolution (e.g., Green et al., 2001).

The detection of spectral features originating in the very proximity of the SMBH is somehow surprising: QSOs are typically very bright, having a bolometric luminosity of the order of 10^{46} erg s $^{-1}$, and most of the energy they radiate consists in strong UV and X-ray emission from the inner regions; such emission is therefore expected to overionize the gas it interacts with at small distances and, as a consequence, no spectral lines should be observed at all. Thus, any model attempting to describe the origin and effect of QSO winds must provide a solution for the so-called overionization problem and explain why gas is not fully ionized and we do see features in QSO spectra. The presence of a shielding material located between the central source of radiation and the outflowing winds could possibly account for the lines we observe.

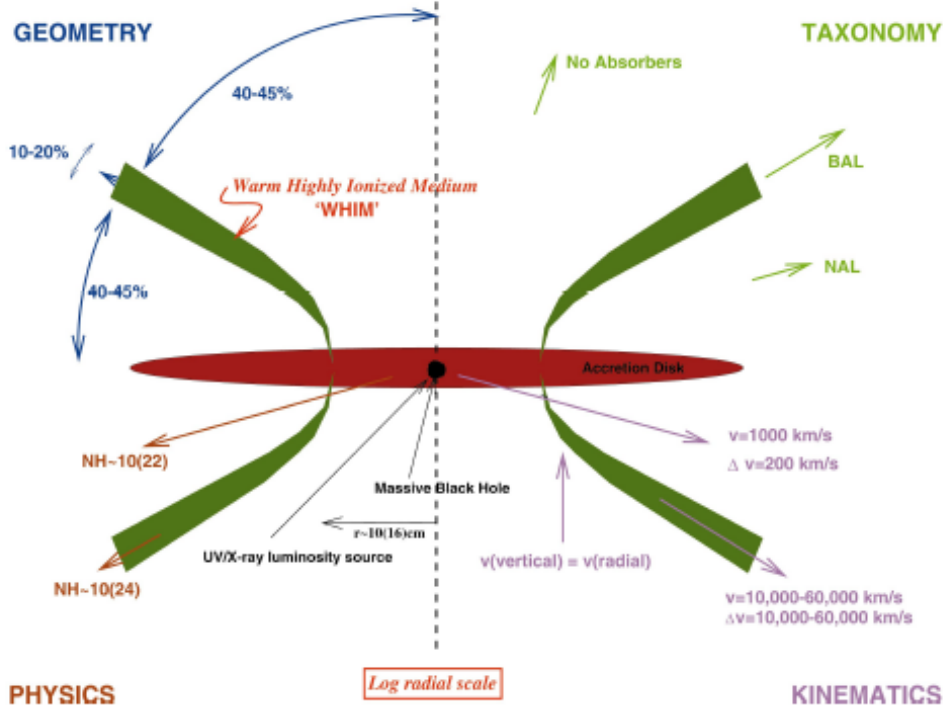


Figure 1.7: Schematic view of QSO wind structure. The four quadrants indicate (*clockwise from top left*) the characteristic opening angles of the structure, the spectroscopic appearance to different observers depending on the viewing angle, the typical outflow velocities along different lines of sight, and characteristic radii and column densities. Credit: Elvis (2000).

No BALs were detected in radio-loud sources until the '90s (e.g., Becker et al., 1997a,b): this was likely due to the shallowness of radio surveys providing QSO spectra where BALs were searched, together with the limited size of the samples where radio emission from BAL QSOs was investigated (e.g., Brotherton et al., 1998). Several additional studies (e.g., Brotherton et al., 2005; Filiz Ak et al., 2012) show that BAL QSOs can be radio-quiet as well as radio-loud sources, hence a number of works (e.g., Rochais et al., 2014, and references therein) nowadays point towards the unification of the two classes.

It is common practice to characterize absorption lines in terms of velocity: the two ends of a line define the maximum velocity v_{\max} and the minimum velocity¹² v_{\min} ; this allows to define the central velocity $v_c = (v_{\max} + v_{\min})/2$, which is an indicator of the position of the feature, and the velocity difference $\Delta v = |v_{\max} - v_{\min}|$, which quantifies the width of the absorption line.

Traditionally, absorption lines are classified on the basis of their width Δv in velocity units: we will have BALs if $\Delta v \geq 2000 \text{ km s}^{-1}$, mini-BALs when $500 \leq \Delta v < 2000 \text{ km s}^{-1}$, and NALs for $\Delta v < 500 \text{ km s}^{-1}$. In turn, BALs are classified in three groups on the basis of the observed transitions: HiBALs exhibit only absorption features from highly-ionized species, such as C IV, Si IV, N V; LoBALs are characterized by spectra marked by lower-ionization lines, such as Al II, Al III, Mg II, in addition to the mentioned high-ionization absorption features; when iron lines, such as Fe II, Fe III, are observed in addition to both high- and low-ionization lines, we will have FeLoBALs (e.g., Voit et al., 1993; Hazard et al., 1987).

In the last decades BALs have been extensively studied with the aim of gaining insight into the geometry and physics of QSOs, the associated winds, and the emission/absorption processes that characterize them. A number of works have shown that 15% of BAL troughs in QSO spectra lie at least 10% below the continuum (e.g., Weymann et al., 1991; Gibson et al., 2009; Trump et al., 2006, and references therein); though such troughs are generally blueshifted because of the high speed of the outflowing gas, examples of QSO spectra exhibiting redshifted absorption features can be found in the literature (e.g., Hall et al., 2013). Figure 1.8 shows a window in the SDSS spectrum of a QSO where two BALs can be observed, highlighted

¹²The definition of v_{\max} and v_{\min} does not take into account the minus sign, due to the absorption line blueshifting.

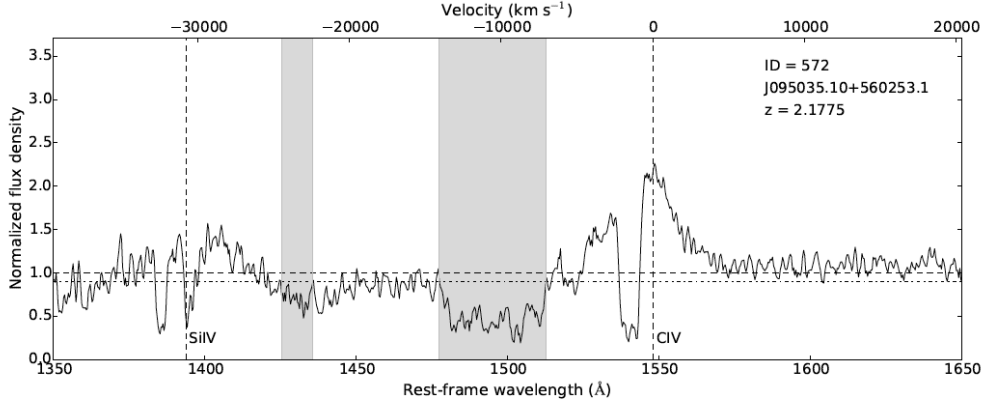


Figure 1.8: Example of SDSS spectrum restricted to the wavelength range where C IV BALs can be observed. On top right of the panel we report the SDSS ID and the redshift of the source. Plots are restricted to the wavelength range $1350 - 1650 \text{ \AA}$ for a better visualization of the window where C IV BALs can be observed. Rest-frame wavelength (bottom) and velocity (top) are reported on the horizontal axes, while the normalized flux density is shown on the vertical axis. The horizontal dashed line represents the level where normalized flux density equals unity, while the dash-dot line indicates the 0.9 level for the normalized flux density: the analysis we are going to describe in Part II of the present thesis concerns BAL troughs extending below this threshold. The two vertical dashed lines represent the rest-frame wavelengths corresponding to the Si IV and C IV emission lines (1394 \AA and 1549 \AA , respectively). The two grey-shaded regions indicate two C IV BAL troughs detected at different velocities.

by the grey-shaded regions.

Since the '80s, several studies have revealed that the equivalent width (EW) of BAL troughs can vary on typical rest-frame timescales ranging months to years (and sometimes the timescale is much shorter; see, e.g., Grier et al. 2015, where variability on rest-frame timescales of ≈ 1.20 days of a C IV BAL trough is discussed).

The most accredited hypotheses about the origin of BAL variability involve changes in the features (e.g., density, geometry) of the absorber, originating from rotations and/or variations in the features of the outflowing

winds (e.g., Proga et al., 2000); this leads to variations in the covering factor (i.e., the fraction of radiation emitted by the central engine and blocked by the absorbing material) along the line of sight, depending on the velocity structure of the outflows. Variations in the absorber could also concern changes in the ionization level and this, in turn, could reflect into absorption-line variations, as the ion abundances vary, causing the weakening/strengthening of the absorption lines (e.g., Barlow, 1993). The investigation of BAL variability thus returns information about the structure and dynamics of the outflowing winds, hence constraining QSO outflows.

Past studies about BAL variability were generally limited by the difficulty in obtaining multi-year observations for large samples of sources, hence typically either the sample size or the observing baseline length were given up. For instance, in Barlow (1993) a sample of 23 QSOs is monitored by means of observations covering a ≈ 1 yr timescale, and variability is detected in 15 of the sources; in Lundgren et al. (2007) the QSO sample consists of 29 units and C IV BAL variability is investigated over a < 1 yr baseline, while Gibson et al. (2008) characterize C IV BAL variability in 13 QSOs observed over 3 – 6 yr, depending on the source. Some works report disappearance (e.g., Lundgren et al., 2007) or emergence (e.g., Capellupo et al., 2012) instances in individual sources.

In Filiz Ak et al. (2012) a statistical analysis of C IV BAL disappearance is presented for the first time: the study is an ancillary project making use of data from different surveys that are part of the Sloan Digital Sky Survey-I/II/III (SDSS; York et al. 2000), and is based on a sample consisting of 582 QSOs with observing baselines ranging 1.1 – 3.9 yr and with spectral coverage from at least two surveys, so that the investigation of BAL variability by means of the comparison of at least two spectra of the same source is possible.

The analysis we are going to present in the second part of the present work takes the cue from Filiz Ak et al. (2012) and takes advantage of observations from the Baryon Oscillation Spectroscopic Survey (BOSS; e.g., Dawson et al. 2013 that were not available at the time the latter work was published, thus significantly extending the sample of inspected sources, which turns out to be more than doubled. This is the first time that a sample this large is used for BAL disappearance analysis.

**Variability-selected Active
Galactic Nuclei in the VST
Surveys of the COSMOS Field
and the Chandra Deep Field
South**

Chapter 2

The VST Multi-epoch Survey of the COSMOS Field

The VLT Survey Telescope (VST; e.g., Capaccioli & Schipani 2011) is the result of a collaboration between the European Southern Observatory (ESO) and the INAF-Osservatorio Astronomico di Capodimonte, Napoli, Italy. It is housed at Cerro Paranal, in the Chilean Atacama desert, in the proximity of the four units constituting the Very Large Telescope (VLT), for which it was designed as a support facility: the VST is in fact conceived to support the VLT through multi-color wide-field imaging surveys and through the identification and preliminary study of potentially interesting targets, to be observed in greater detail by the more sensitive VLT telescopes. It is one of the largest ground-based telescopes dedicated to surveys in the optical waveband, extending from 0.3 to 1.0 μm , to the UV and the NIR.

2.1 The VST and the SUDARE Survey

The telescope has a modified Ritchey-Crétien configuration and an alt-azimuth mount; it is equipped with a 2.650 m-diameter primary mirror, a 0.938 m-diameter secondary mirror, and a single focal plane detector, OmegaCAM (Kuijken, 2011). This consists of a mosaic of 32 CCD cameras which correspond to a total of 268 million $15\ \mu\text{m}$ -size pixels over a $26 \times 26\ \text{cm}^2$ area; the focal plane scale is $0.214''/\text{pixel}$, returning a FoV of $1^\circ \times 1^\circ$. The 32 CCD cameras are flanked by four additional CCDs: two of them are intended for auto-guiding, hence their task is to accurately track field position and rotation; the other two are mounted at a distance of 2 mm from the focal plane—one in front of it, the other behind—and record out of focus star images for curvature wavefront sensing and control of the active optic system of the telescope. Such a detector returns ≈ 50 Gbytes of raw data per observing night.

The telescope is provided with a selection of 12 filters including the SDSS *ugriz* set, the Johnson *B* and *V* filters, a number of narrow-band filter mosaics, a Strömgren *v* filter, and a special calibration filter. It is dedicated to several astrophysical and cosmological topics, encompassing Solar System studies as well as extrasolar planet search, the investigation of the structure and evolution of the Galaxy as well as the survey of the extragalactic Universe, aimed at a deeper knowledge of its early history; VST data are also employed in projects meant to refine the calibration of the cosmic distance ladder and in the search for active galactic nuclei and cosmic structures at medium-high redshift.

The VST is presently involved in a variety of Galactic and extragalactic survey programs; in the work presented here, we make use of data from the SUPernova Diversity And Rate Evolution survey (SUDARE; Botticella et al.

2013; Cappellaro et al. 2015; Botticella et al. 2016). The project aims at analyzing the trend of the rates of different SN types in the redshift range $0.3 < z < 0.8$, together with possible correlations with the properties of the host galaxies and any dependence of such correlations on redshift and/or on the stellar population the SNe belong to. The survey concerns two distinct sky regions centered on the COSMOS field¹ and the Chandra Deep Field South²; observations are available in the g , r , and i band and, in the case of the CDFS, additional observations in the u band are provided by the VST Optical Imaging of the CDFS and ES1³ (VOICE; Vaccari et al. in prep.) Field Survey: altogether, SUDARE and VOICE survey two 4 square degree areas, with the goal of providing optical data for such regions, also surveyed by several other observatories (e.g., *Herschel*/HerMES, *Spitzer*-Warm/SERVS, VISTA/VIDEO) at different wavelengths. The combined use of the available datasets ensures a reliable estimate of photometric redshifts and the definition of constraints on stellar masses for more than one million galaxies in the redshift range $1 < z < 5$, thus allowing to investigate the relation among star formation, stellar mass and cosmic environment, in the general context of galaxy evolution.

¹The COSMOS field is named after the Cosmic Evolution Survey, dedicated to the study of galaxy formation and evolution and the large scale structure of the Universe; it is a 2 square degree area centered on RA= 10h 00m 28.6s, Dec = +02° 12' 21.0". We will deal with this region more extensively in the following.

²The Chandra Deep Field South is a sky region covering a 0.11 square degree area and centered on RA= 3h 32m 28.0s, Dec = -27° 48' 30", named after the *Chandra X-ray Observatory*, which performed deep surveys of the area.

³The *Infrared Space Observatory* (ISO) first observed several areas of the sky at high galactic latitudes, covering about 20 square degrees in total; ES1, or ELAIS1, stands for European Large Area ISO Survey.

2.2 The SUDARE-COSMOS Dataset

We made use of the SUDARE observations to lead a project aimed at identifying AGNs on the basis of their optical variability. A first analysis was performed on the COSMOS field, which is one of the most widely surveyed areas in the sky, observed by most of the ground-based and space-based telescopes over the entire electromagnetic spectrum: this implies that a large set of diagnostics at different wavelengths are available for AGN selection and for the validation of our results. The method was then validated by the analysis performed on the CDFS, whose results are illustrated in Falocco et al. (2015) and Poulain et al. (in prep.).

The sampling time of SUDARE is of three days for r band observations and ten days for all the other bands, depending on various observational constraints. In Section 1.3 we mentioned that there are several works in the literature (e.g., Cristiani et al., 1997; Di Clemente et al., 1996; Schmidt et al., 2012) dealing with the dependence of AGN variability on the optical waveband, and showing that variability is larger for shorter rest-frame wavelengths; hence, in principle, g band data would have been the best choice for our analysis, as they would have allowed us to obtain a larger sample of optically variable AGN candidates; nevertheless, given the higher temporal sampling, we chose to use r band data for our analysis. Anyway, we estimated that the variability increase would have been on the order of just 30% with respect to the r band, so one should not expect significant improvement if using g band data. Conversely, i band variability would have been characterized by a smaller amplitude if compared to the r band, hence the sample would have probably been smaller; in addition, i band images are generally affected by sky-emission fringes, hence their photometry is typically worse than in r band images.

The VST imaged the COSMOS field during three different observing programs: from December 2011 to May 2012, then again from December 2013 to April 2014, and once more from December 2014 to March 2015. The final dataset consists of 65 epochs, and full information about each of them is reported in Table 2.1; each epoch is obtained from a number of dithered exposures (usually five, corresponding to a total exposure time of 1800 s, but see Table 2.1 for exceptions) of the imaged field corresponding to the same observing block (OB)⁴. Dithering consists in a small shifting in the pointing of the telescope between different exposures, and it is used to obtain an image with a higher signal-to-noise ratio (S/N) than in individual exposures: because of the shift, each object is imaged in a slightly different position on the detector in each exposure, and this allows to identify and consequently remove possible artifacts, such as cosmic rays or satellite tracks. Our dataset also includes a weight map⁵ for each epoch. Limited to the first observing program (December 2011 - May 2012), we also have a stacked image, obtained as the median of all the exposures from the first observing program having a seeing FWHM $< 0.80''$, corresponding to a total exposure time of 19800 s. The stacked image has a limiting magnitude $r(\text{AB}) \approx 26$ mag at $\sim 5\sigma$ above the background rms, while single epochs are generally characterized by $r(\text{AB}) \lesssim 24.6$ mag.

The COSMOS field as imaged by the VST is shown in Figure 2.1.

⁴Throughout the present work we will keep the distinction between epochs and exposures, an epoch being the combination of different exposures, as explained above.

⁵A weight map consists of an image having the same size of the corresponding epoch, and accounts for the different noise levels in different pixels by associating a weight $w_i = 1/s_i^2$ to each pixel i , s_i^2 being the noise variance corresponding to the pixel.

2.2.1 The *VST-Tube* Pipeline

The single exposures constituting our dataset had already been reduced and combined in order to obtain epochs by means of the *VST-Tube* pipeline (Grado et al., 2012), designed to process the large amount of data returned by the VST. Essentially, the pipeline takes care of overscan correction, bias subtraction, and flat-field correction, as well as CCD gain harmonization and illumination correction (Capaccioli et al., 2015). Astrometry correction and photometric calibration follow, then single exposures are combined together.

The overscan correction consists in subtracting the median value measured over an overscan region in the image; a sigma-clipped average of the bias frames then returns the master-bias frame, which is subtracted to each exposure. In order to obtain the master-flat, a master twilight flat and a super sky-flat are combined: the first allows to correct for the different sensitivity in different pixels, while the latter takes into account variations in the low spatial frequency gain; both are obtained by means of a robust sigma-clipped average of twilight frames and science frames, respectively, which are to be overscan- and bias-corrected before.

Gain harmonization is necessary in order to have the same zero-point throughout the CCD cameras making up the detector: this is achieved if the background level differences in adjacent CCDs are minimal, so proper CCD gain coefficients are to be determined. Illumination correction is required as images from wide-field telescopes are characterized by a spatially-varying background component due to scattered light; this causes the photometric response of the detector to be position-dependent (Andersen et al., 1995). In order to correct for such an effect, an illumination correction map is needed: this is obtained by means of a proper selection of standard fields for which a reference catalog of stars (in this case, the catalog from Aihara et al. 2011,

from the Data Release 8 of the SDSS) is available, so that the instrumental magnitudes of specific stars in such fields can be compared to the magnitudes of the same sources reported in the catalog. A generalized additive model (Wood et al., 2011) is used to fit the magnitude difference as a function of the position on the detector. In order to get a uniform background, the illumination correction map is then properly rescaled and subtracted from the images.

The absolute photometric calibration is performed on a chosen epoch, by means of a comparison of the instrumental magnitudes to reference ones, in selected sky fields; the minimization of the quadratic sum of magnitude differences between overlapping exposures then allows to perform the relative photometric correction. Images are finally resampled and then combined as a weighted average making use of the tool *SWARP* (Bertin et al., 2002).

Table 2.1: COSMOS dataset. Epoch number, OB identification number, date, and seeing FWHM for the 65 epochs; epochs are listed in chronological order. The three epochs marked by two stars were excluded from the analysis at the very beginning of it (and this is why we never assigned them an identifying number), as the exposures turned out to be affected by severe aesthetic artifacts and defects; all the epochs marked by a single star were excluded in a later step of the analysis (see Chapter 3). Epochs 52 and 53 were obtained by the combination of ten exposures, instead of five, for a total exposure time of 3600 s; for all the remaining epochs, five exposures were combined together, and the total exposure time is 1800 s.

epoch	OB-ID	obs. date	seeing (FWHM) (arcsec)	epoch	OB-ID	obs. date	seeing (FWHM) (arcsec)
1	611279	2011-Dec-18	0.64	32	986626	2014-Jan-12	0.73
2	611283	2011-Dec-22	0.94	33*	986630	2014-Jan-21	1.18
3	611287	2011-Dec-27	1.04	34	986633	2014-Jan-24	0.80
4	611291	2011-Dec-31	1.15	35*	986636	2014-Jan-27	1.09
5	611295	2012-Jan-02	0.67	36*	986642	2014-Feb-02	1.21
6*	611299	2012-Jan-06	0.58	37*	986648	2014-Feb-09	1.28
7*	611311	2012-Jan-18	0.62	38	986652	2014-Feb-19	0.89
8	611315	2012-Jan-20	0.88	39	986655	2014-Feb-21	0.93
9	611319	2012-Jan-22	0.81	40	986658	2014-Feb-23	0.81
10*	611323	2012-Jan-24	0.67	41	986661	2014-Feb-26	0.81
11	611327	2012-Jan-27	0.98	42	986664	2014-Feb-28	0.77
12	611331	2012-Jan-29	0.86	43*	986667	2014-Mar-04	1.11
13	611335	2012-Feb-02	0.86	44	986670	2014-Mar-08	0.91
14*	611351	2012-Feb-16	0.50	45	986674	2014-Mar-21	0.96
15	611355	2012-Feb-19	0.99	46	986677	2014-Mar-23	0.92
16	611359	2012-Feb-21	0.79	47*	986680	2014-Mar-25	0.66
17	611363	2012-Feb-23	0.73	48	1095777	2014-Mar-29	0.89
18	611367	2012-Feb-26	0.83	49	1095783	2014-Apr-04	0.58
19	611371	2012-Feb-29	0.90	50	986683	2014-Apr-07	0.61
20	611375	2012-Mar-03	0.97	51	1136410	2014-Dec-03	1.00
20b**	611379	2012-Mar-06	0.82	52*	1136444	2014-Dec-30	1.04
21	611387	2012-Mar-13	0.70	52b**	1136453	2015-Jan-04	0.77
22	611391	2012-Mar-15	1.08	53	1136457	2015-Jan-10	0.71
23	611395	2012-Mar-17	0.91	54*	1136478	2015-Jan-25	1.30
24	768813	2012-May-08	0.74	55	1136481	2015-Jan-28	0.90
25	768817	2012-May-11	0.85	56	1136490	2015-Jan-31	0.73
26	768820	2012-May-17	0.77	57	1136503	2015-Feb-15	0.70
27*	768826	2012-May-24	1.27	58*	1136506	2015-Feb-19	1.14
27b**	986608	2013-Dec-25	0.88	59*	1136518	2015-Mar-01	1.33
28	986611	2013-Dec-27	0.72	60	1136531	2015-Mar-10	0.80
29	986614	2013-Dec-30	1.00	61	1136540	2015-Mar-14	0.84
30	986617	2014-Jan-03	0.86	62	1136543	2015-Mar-19	1.00
31	986620	2014-Jan-05	0.81	stacked	-	-	0.67

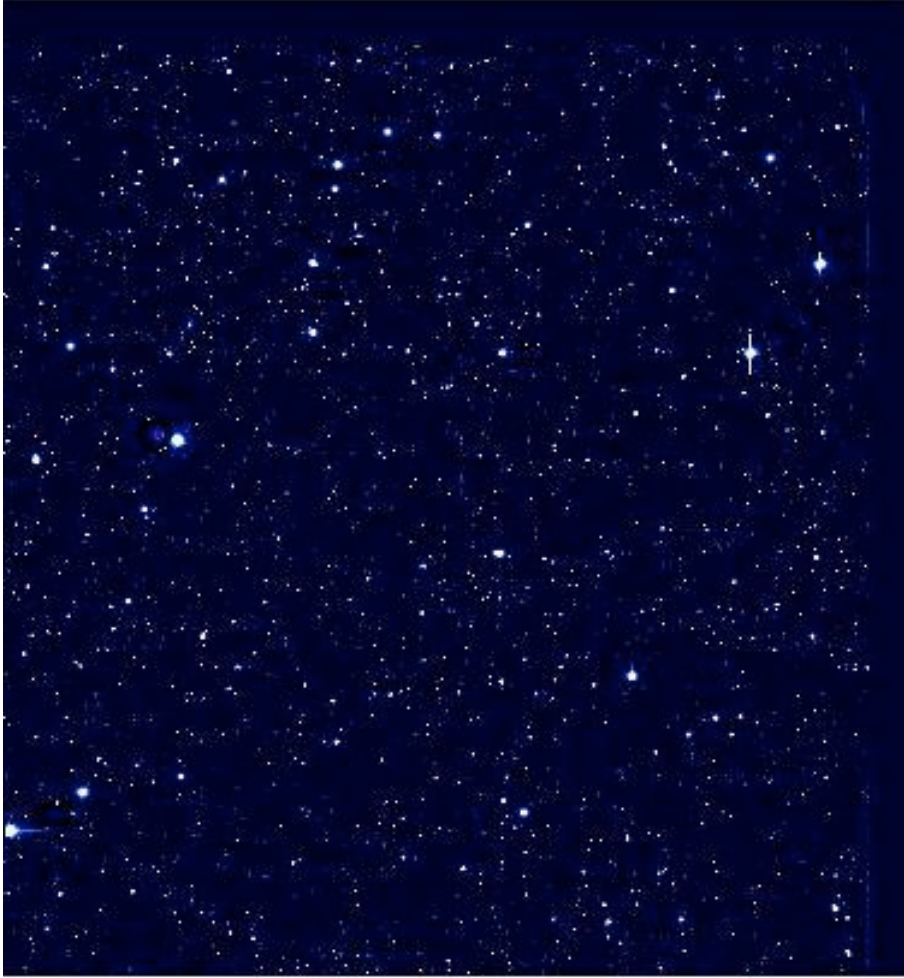


Figure 2.1: Our stacked image ($\approx 16\text{K} \times 16\text{K}$ px, corresponding to $\approx 1^\circ \times 1^\circ$), showing the VST COSMOS field; it was obtained as the median of the fifty-seven exposures from the first COSMOS observing program having a seeing value $< 0.80''$.

Chapter 3

Selection of AGN Candidates

The effort to search and characterize AGNs through optical variability started at the beginning of the SUDARE survey. As the survey progressed, we incrementally selected new AGN samples taking advantage of the increasing temporal baseline. A first analysis of optical variability-selected AGNs in the SUDARE-COSMOS field was presented in De Cicco et al. (2015) and concerns observations spanning a five month temporal baseline.

At the time the work was started, the available epochs were those from the first observing program (December 2011 - May 2012), corresponding to epochs from 1 to 27 in Table 2.1. At present, the analysis has been extended to a set of 47 epochs out of 65 (see Table 2.1); the reason why the remaining epochs were not taken into account will be explained in Section 3.1.

In the following we present the analysis of the AGN candidates selected over the full baseline from December 2011 to March 2015; differences with respect to the approach followed in De Cicco et al. (2015), as well as a comparison with the results obtained in the two cases, will be extensively discussed. Hereafter we will refer to the 47 epoch dataset as $D1$, while the subsample of 27 epochs corresponding to the first observing program and

analyzed in De Cicco et al. (2015) will be labeled *D0*.

The method adopted to identify optically variable sources and hence define our sample of AGN candidates traces the approach proposed by Trevese et al. (2008). The present work aims at pushing the AGN selection based on variability towards deeper limits than in past works, in order to test the strength of the method in the perspective of future wide-field surveys, such as the Large Synoptic Survey Telescope (LSST; e.g., Abell & the LSST collaboration 2009).

We implemented the same method to extend the search for optically variable AGNs in the CDFS, making use of additional diagnostics to confirm our sources, and further validating the effectiveness of the technique. We shall go back to this in Chapter 5; full results are described in Falocco et al. (2015) and Poulain et al. (in prep.).

3.1 Catalog Extraction

For each epoch we obtained a catalog of sources by running *SExtractor* (Bertin & Arnouts, 1996), a software developed to identify astronomical objects, for which it returns a set of parameters and physical quantities of interest for photometric studies.

As a first step, *SExtractor* scans the astronomical image and separates objects from background: on the basis of a set flux threshold, pixels are considered to belong either to a source, or to the background. Weight images –when available– allow to assign each pixel a statistical weight. The detection of fainter objects can be eased by applying a smoothing filter to the image. A background map and the corresponding rms map are built, then the background is subtracted from the image. *SExtractor* then performs as-

trometric, photometric, and geometric measurements returning information about position, brightness and a number of parameters for each source in the catalog.

The number of sources detected in each image, and thus constituting each catalog, approximately ranges 35000 – 55000, being higher for epochs with lower seeing values. A bad seeing reduces the number of detected objects in an image as it makes object deblending difficult: the larger the images of two neighboring objects, the higher the risk of overlapping and light contamination. For each source, the catalog provides positional coordinates, magnitudes in a set of apertures (see Section 3.2), and a stellarity index. This is an object morphological classifier based on a neural network output; basic input parameters for the neural network are the pixel scale and the seeing, together with other parameters the classification indirectly depends on. The index ranges from 0 to 1, going from extended objects (e.g. bright galaxies) to unresolved sources (e.g., faint galaxies, stars, and QSOs). For each source, a measure of the half-light radius is also available: such radius defines a circle which encloses approximately half the total light from a point-like source.

The catalogs of sources obtained for each epoch were matched together within a matching radius of $1''$, thus returning the light curves of all the detected objects. The master catalog thus obtained contained 98754 sources; part of them were detected just in one or few epochs and, in most of the cases, are fake objects and have been removed from our sample at a later step (see Section 3.4).

In order to test the quality of each epoch in our dataset, we visually inspected each of them as a first step; in some epochs, several sources have an unusual irregular shape, as shown in the example presented in Figure

3.1; the inspection of the corresponding weight map revealed that the pixels corresponding to such objects had been assigned infinite variance due to an incorrect cosmic ray removal during the image reduction process. In other cases, other defects were apparent; for instance, lines tracing the CCD contours were visible, as a result of a bad image reduction. We decided to exclude 15 epochs where such defects were detected; this may be quite a conservative choice in some cases but, given the long baseline and the high sampling frequency, the exclusion of some epochs does not have a dramatic impact on the dataset, and allowed us to focus our analysis on the best quality images. Three additional epochs, labeled as 20b, 27b, and 52b in Table 2.1, have been excluded from the start because of severe aesthetic artifacts and defects¹; this leaves 47 epochs constituting our *D1* dataset.

3.2 Aperture Selection and Correction

In order to quantify the variability of a source, it is necessary to get its light curve, hence a suitable aperture to properly measure magnitudes in each epoch is required. Since our main interest was AGN identification, the optimal choice would have been an aperture allowing to maximize the collection of flux from a possible active nucleus while minimizing the contamination

¹The defects in epoch 20b are mainly arcs originating from reflections internal to the telescope due to the presence of nearby bright stars, while epochs 27b and 52b are characterized by an unusually high noise: this was caused by a very strong and non-homogeneous illumination, originating in turn from moonlight reflections by clouds; moreover, there were some tracking problems with the telescope. As a consequence, there was nothing we could do to fix the three epochs in order to include them in our analysis; nevertheless, given the large number of available epochs, adding three more which would not even extend the baseline was not a crucial issue.

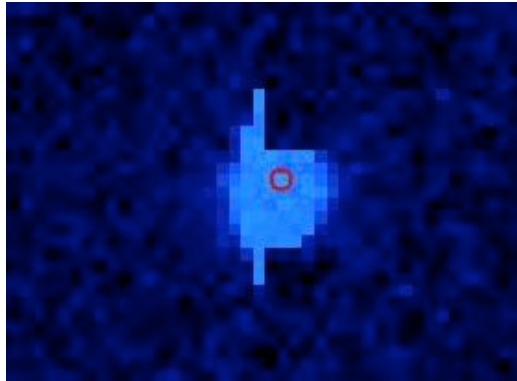


Figure 3.1: Example of source with a weird shape detected in one of the epochs which turned out to be affected by bad cosmic ray removal. The red circle pinpoints the position of the source according to the coordinates reported in the corresponding catalog.

from the host galaxy. We identified such aperture starting from the average stellar half-light radius of the point-like sources in each epoch as, for objects having self-similar profiles –such as stars– this value should be roughly constant due to atmospheric seeing. The average value of the half-light radius is of course a more reliable indicator than individual radii as, in case of very bright or faint individual objects, the corresponding radius could be affected by saturation or very low S/N, respectively. Since active nuclei are typically point-like, an aperture radius which is twice the stellar half-light radius constitutes a suitable choice, as it will enclose most of the light from stellar-like sources. Each of the catalogs obtained by means of *SExtractor* provides, among other parameters, the half-light radius for each source.

In order to estimate the average stellar half-light radius for each epoch, we extracted a temporary set of catalogs including magnitudes computed in a set of apertures common to all epochs and ranging 4 – 115 px, corresponding to $1''$ – $25''$; such apertures are the same used by the Cambridge Astronomical Survey Unit (CASU) in the analysis of data from the Visible and Infrared Survey Telescope for Astronomy (VISTA; e.g., Sutherland et

al. 2015), which observes in a partially overlapping wavelength range as VST and, being located in the proximity of it, operates in similar atmospheric conditions.

For each epoch, the stellar half-light radius turned out to be $\approx 2 - 2.5$ pixels, corresponding to $\approx 0.43'' - 0.54''$; as a consequence, a $2''$ -diameter aperture looked like a suitable choice for our analysis.

An aperture correction accounting for the epoch-to-epoch seeing differences was necessary. Essentially, the correction consists in applying a corrective factor to each epoch, so that the fraction of collected light is independent on the seeing. A $2''$ -diameter aperture generally encloses 70% of the flux from a source when centered on the source centroid; we set the fraction of collected light to be the 90% of the total flux (but this is an arbitrary choice, as long as the fraction is the same for every epoch); hence, in order to derive the aperture correction factor for each epoch, we made use of the growth curves of bright stars across the FoV, chosen as a reference: these were required to be detected in every epoch and to be non-saturated, isolated from other sources, and distant from the edges of the field and from regions affected by defects such as bright-star halos or reflections. For each epoch we determined the ratio of the flux from the reference star enclosed in a $2''$ -diameter aperture to the flux corresponding to 90% of the total; the obtained ratio was assumed to be the corrective factor for that specific epoch. In Figure 3.2 we show the flux growth curve as a function of aperture for one of our reference stars, measured in epoch 1.

Such a technique for aperture correction is tailor-made for point-like sources but, in the case of fainter AGNs, there is the risk of a significant light contribution from the host galaxy, within a $2''$ -diameter aperture, that would be erroneously attributed to the active nucleus and would hence be

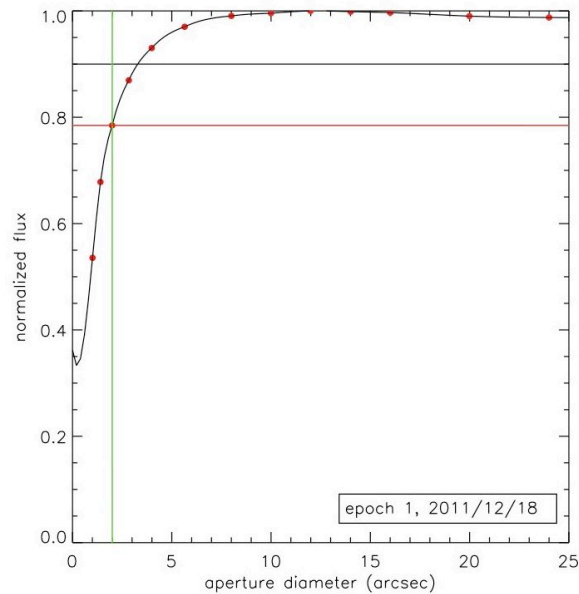


Figure 3.2: Flux growth curve as a function of aperture for one of the chosen reference stars. The difference between 90% of normalized flux (black line) and the flux into a $2''$ -diameter aperture (red line) is shown; the green line indicates the $2''$ aperture, corresponding to 9.5 px. Data refer to the epoch 1.

improperly taken into account in our correction process, leading to an overestimation of the source variability. We will go back to this point in Section 3.4, after describing of our selection criterion, and will explain why the possible contamination from the host galaxy in case of extended objects constitutes a minor concern in the framework of our analysis.

3.3 Defects and Aesthetic Artifacts

Observations of the COSMOS field from the first observing program correspond to the very beginning of the VST activity, hence the various epochs turned out to be affected by a number of defects in the detector electronics, as well as aesthetic artifacts; VST data-users, including us, had to deal with

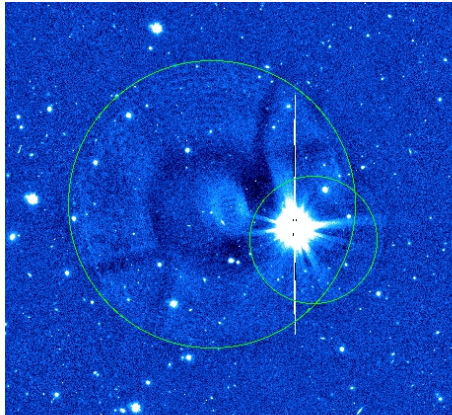


Figure 3.3: A saturated star generating two halos (roughly delimited by the two green circles) in its proximity; their diameters are $\approx 350''$ and $\approx 150''$.

them for the first time, and the most natural path to follow resulted in a conservative approach. Most of the problems arose from a poor knowledge of the detector response, and were fixed in the following months. In particular, the most relevant problem was a CCD characterized by random variations in its gain factor, reflecting into a huge concentration of fake sources in the corresponding area of the detector; this led to the exclusion of the image region corresponding to that CCD from each epoch. Similarly, other regions were excluded in some or all the epochs, depending on the circumstances. Examples are the edges of each epoch (corresponding to a $\approx 2' - 4'$ width), characterized by a very low S/N, satellite tracks, regions affected by reflections of scattered light, and –last but not least– bright star halos. The last ones are a common feature in wide field surveys, and constitutes what we mainly needed to exclude by applying suitable masks to the images at issue. Figure 3.3 shows a region to be masked as an instance.

A mask is a file where positional information about the regions to exclude is stored. Masks were shaped making use of the procedure developed by Huang et al. (2011). Each catalog was to be combined with the corre-

sponding mask; we used the same basic mask for all epochs, adding additional regions to exclude, depending on the epoch. The masking process reduced the number of objects in each catalog by several thousands. The catalogs where the lowest fraction of objects were masked correspond to the epochs having the best seeing value (among the ones that we did not exclude from our analysis), and the fraction is $\approx 21\%$; conversely, in the epochs with highest seeing values we masked up to 25% of the sources. The percentages we found are consistent with the fractions we obtained from the analysis limited to *D0* data.

3.4 The Sample of AGN Candidates

When dealing with *D0* data, we chose to limit our variability analysis to the subsample of objects detected in at least six epochs (corresponding to 20% of the dataset), in order to reduce contamination from fast transients and/or spurious sources; this is an arbitrary threshold, nonetheless we found out that the inclusion/exclusion of a couple epochs returned largely overlapping samples of sources, with a few percent difference, and that the difference in the corresponding samples of AGN candidates was limited to a few units. On the basis of such tests, and since a source detected in ≥ 6 epochs is likely not spurious (at least when defected areas are properly masked), we kept this same threshold when dealing with *D1* data. In our work analyzing *D0* data we focused on sources with $r(\text{AB}) \leq 23$ mag, as such a threshold approximately corresponded to the completeness limit of our single epoch catalogs; here, in order to be consistent and be able to properly compare results obtained from the analyses over different baselines, we adopted the same magnitude threshold; nonetheless, we point out that the completeness

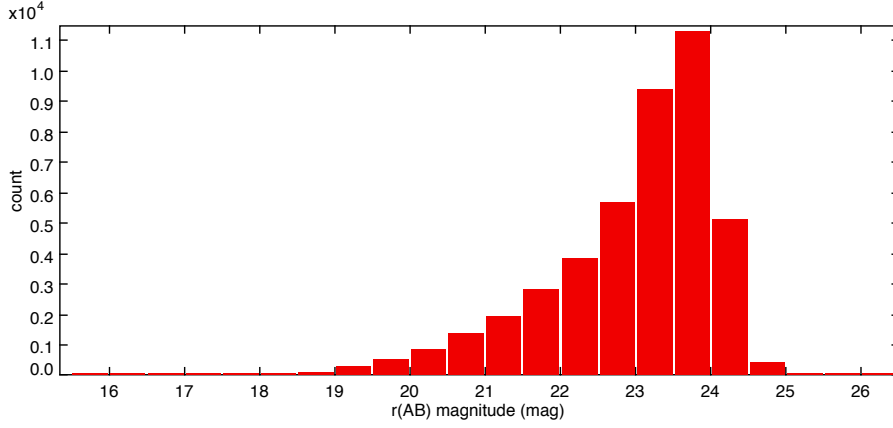


Figure 3.4: Magnitude distribution for all the sources detected in epoch 49.

limit for the epochs obtained in observing programs more recent than the first one is $r(\text{AB}) \lesssim 24$ mag, hence we could have extended our analysis to sources in the magnitude bin $23 - 24$ mag, and this is one of the things we plan to do before publishing the new results. In Figure 3.4 we show the magnitude distribution for the sources in epoch 49, i.e., the one with the best seeing value; 87% of the sources have $r(\text{AB}) \leq 24$ mag.

The choice of a magnitude threshold may result in the exclusion of a number of fainter AGNs but, on the other hand, below this limit we may be missing source detection in some epochs due to completeness and this can lead to spurious variability. The applied selection criteria returned a sample of 19226 sources (hereafter, main sample); we remind that this includes both variable and non variable sources.

In order to define a variability threshold, we computed the average magnitude $\langle \text{mag}_i^{lc} \rangle$ and the corresponding rms deviation σ_i^{lc} for each source i from the corresponding light curve:

$$\langle \text{mag}_i^{lc} \rangle = \frac{1}{N_{epo}} \sum_{j=1}^{N_{epo}} \text{mag}_i^j, \quad \sigma_i^{lc} = \left[\frac{1}{N_{epo}} \sum_{j=1}^{N_{epo}} (\text{mag}_i^j - \langle \text{mag}_i^{lc} \rangle)^2 \right]^{\frac{1}{2}}, \quad (3.1)$$

where the superscript lc stands for “light curve”, j runs through the epochs, and N_{epo} is the number of epochs where we detect the source i . We then computed the running average $\langle \sigma_i^{lc} \rangle$ of the rms deviation, and the corresponding rms deviation $\text{rms}_{\langle \sigma_i^{lc} \rangle}$, over 0.5 mag-wide bins. This allowed us to define a variability threshold

$$\sigma_i^{lc} \geq \langle \sigma_i^{lc} \rangle + 3 \times \text{rms}_{\langle \sigma_i^{lc} \rangle} , \quad (3.2)$$

and the variability significance σ_i^* for the i -th source, following Bershadsky et al. (1998):

$$\sigma_i^* = \frac{\sigma_i^{lc} - \langle \sigma_i^{lc} \rangle}{\text{rms}_{\langle \sigma_i^{lc} \rangle}} . \quad (3.3)$$

Equations 3.2 and 3.3 imply that, in order to be classified as variable, our sources must have a significance $\sigma^* \geq 3$. The sources above the variability threshold turned out to be 464 (2.4% of the main sample), and they constitute our temporary sample of AGN candidates. Figure 3.5 shows the variability threshold and the sample of AGN candidates in the plane of σ^{lc} vs. $\langle \text{mag}_i^{lc} \rangle$.

In Section 3.2 we stated that the technique we used for aperture correction fits point-like sources, but implies the risk of contamination from the host galaxy light when dealing with extended objects. In order to quantify such spurious variability increase, we performed the AGN candidate selection described above on different subsamples of sources selected on the basis of their stellarity index (which we introduced in Section 3.1). The stellarity index allowed us to select from our main sample the two subsamples of point-like objects (stellarity index ≥ 0.8) and extended sources (stellarity index ≤ 0.2); we thus repeated the AGN candidate selection procedure on each subsample. This showed that, in the case of extended sources, a slight rise in the variability threshold is observed; this reflects an increase in the average variability of the sources in the subsample of $\lesssim 0.01$ mag. Therefore, the overall effect of an improper correction for extended objects would

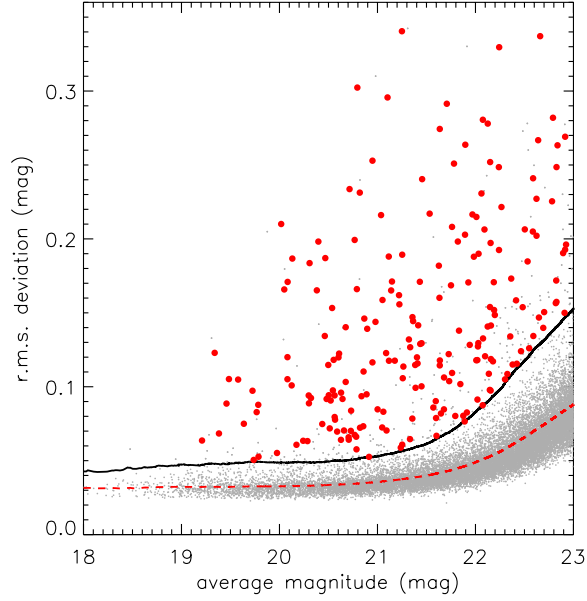


Figure 3.5: Light curve rms σ^{lc} as a function of the average magnitude $\langle \text{mag}^{lc} \rangle$. The dashed line represents the running average of the rms, while the solid line defines the variability threshold. Small grey dots stand for the sources in the main sample; objects above the threshold are assumed to be variable, and large red dots are the candidates that we confirm as AGNs in the following (Chapter 4).

be a more conservative sample of optically variable sources. Moreover, from the stellarity index of the AGN candidates obtained from the selection in each subsample, we see that most of them are compact sources, hence we can conclude that the correction for extended sources is definitely a minor concern.

We expected some of the sources in the temporary sample of AGN candidates to be not intrinsically variable objects: spurious variability can effectively be the result of external factors, such as the non-deblending of a very close companion in the catalog extraction process; in principle, a change in the input extraction parameters could help reducing such occurrences but, in any case, when a source has a very close companion (and, in particular,

when this is a very bright object) it is not possible to tell whether variability is an intrinsic property of the source or it rather arises from epoch-to-epoch point spread function (PSF) variations combined with contamination from the companion. One more class of spurious variable sources consists of very extended and irregular objects, where the identification of the centroid is tricky due to an irregular shape (e.g., late-type galaxies with irregular morphology). Also, there could be sources which happen to fall onto defected areas of the detector (e.g., hot pixels) as well as regions where noise dominates (e.g., close to the edges of the frame), that were not masked properly. We expected the sample to also include a number of contaminants, i.e., intrinsically variable sources that are not AGNs, such as variable stars and transient events such as SNe. The identification and analysis of SNe in the *D1* data is currently ongoing; as for the sample of AGN candidates obtained from the analysis of *D0* data, the contamination by SNe was 14%.

Our temporary sample of AGN candidates included 57 non-deblended sources which were to be excluded from our sample on the basis of what discussed above. This left 407 AGN candidates. In order to identify and reject spurious sources and contaminants, we visually inspected the snapshots of each source in each epoch as well as their light curves, and flagged each source with a quality label ranging 1 to 3 on the basis of the following guidelines:

- (210 sources) strong candidate, no evidence for problems/defects;
- (55 sources) likely variable candidate, potentially affected by the presence of a neighbor or by minor problems;
- (142 sources) very doubtful variability, likely spurious.

We point out that all the objects with a close neighbor within $2''$ (centroid-to-centroid distance) were flagged as 3 if the magnitude difference of the two

sources is < 1.5 , as we assumed that, in such a case, the light from one source strongly affects the other. In Figure 3.6 we present one source per class as an example; the corresponding light curves are also shown.

We decided to reject from our sample of AGN candidates all the sources labeled as 3, corresponding to 31% of the 464 sources constituting the temporary sample of AGN candidates; as a consequence, the variability analysis that we are going to discuss in the following is focused on the remaining 265 sources (hereafter, secure sample). A selection method that leads to the rejection of almost one third of the sample of candidates may seem inefficient, and hence not reliable; nevertheless, we point out that the sample of AGN candidates itself could have been reduced, before sources were flagged, by means of a refinement of the selection algorithm, so as to exclude all the sources with a very close neighbor. This is something we did not focus on at this stage of the work, due to the reasonable size of the sample of candidates to visual inspect; nonetheless, a refinement of the algorithm is necessary if the method is to be used in the analysis of larger samples. Figure 3.7 shows a $1^\circ \times 1^\circ$ view of the COSMOS field and the positions of the sources belonging to the secure sample.

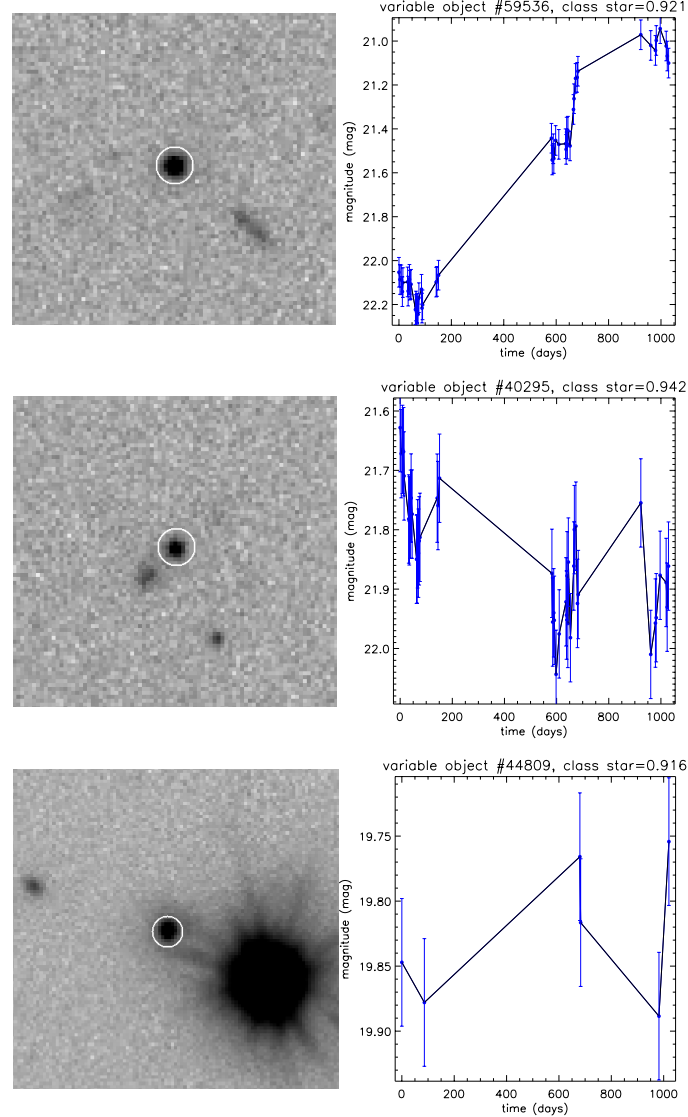


Figure 3.6: Examples of variable candidates assigned to different quality classes, with the corresponding light curves. Images are from epoch 49, i.e., the one having the best seeing value among $D1$ epochs. The white circles correspond to the $2''$ -diameter aperture and are centered on the object average coordinates. Objects labeled 1 (*upper panel*) are generally isolated and free from aesthetic defects. In the case of the objects belonging to class 2 (*middle panel*), potential problems (e.g., the presence of a neighbor) must be taken into account. Objects labeled 3 (*lower panel*) are probably spurious variable sources: the case of a source lying in the halo of an extended and bright neighbor is shown as an example and, from the corresponding light curve, we can see that the source is detected in only six epochs; it is very likely that the variability we detect is a spurious effect of the presence of such a neighbor. The error bars correspond to our adopted threshold of $\langle \sigma_i^{lc} \rangle + 3 \times \text{rms}_{\langle \sigma_i^{lc} \rangle}$.

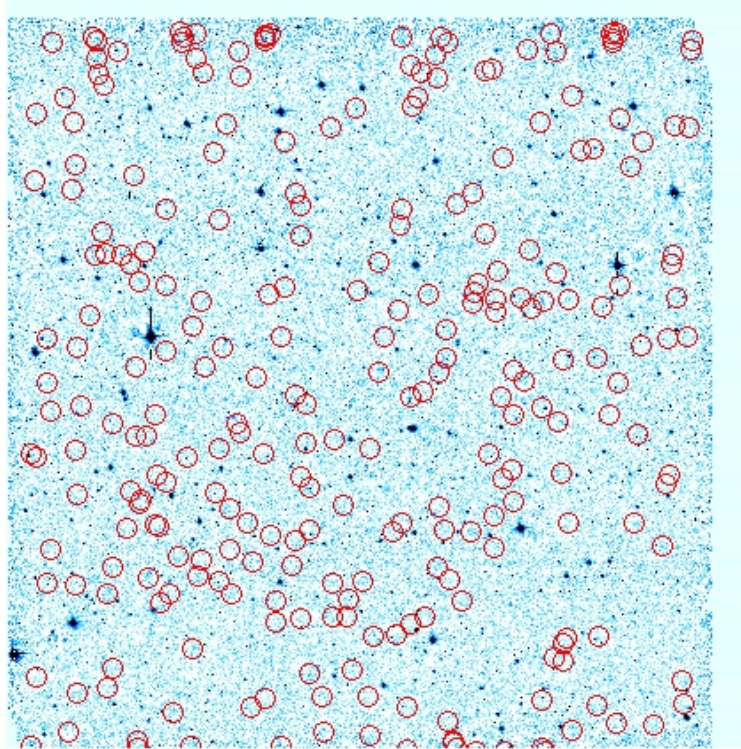


Figure 3.7: VST-COSMOS field and the location of the 265 sources in the secure sample (red circles).

Chapter 4

Validation and Characterization of the Sample of AGN Candidates

In the present chapter we investigate the nature and properties of the sources in our secure sample, aiming at confirming AGNs and at classifying as many of the remaining sources as possible, in order to assess to what extent optical variability is a reliable indicator of the presence of an AGN, and hence a powerful alternative to other more expensive and/or time-consuming techniques (e.g., X-ray identification) usually employed in the search for AGNs.

Due to the widespread literature about the COSMOS field, classification already exists for part of our sample, and it is reported in other COSMOS catalogs; furthermore, several catalogs at different wavelengths are available for most of the sources. This allowed us to validate our results by means of a large amount of ancillary data.

4.1 X-ray Counterparts

X-ray emission, especially if coupled with variability, is the strongest, most reliable, and unambiguous indicator of the presence of an AGN. In the present work we made use of two X-ray catalogs of COSMOS sources:

- the catalog containing optical and NIR counterparts of the sources in the *Chandra*-COSMOS Legacy Catalog (Marchesi et al., 2016; Civano et al., 2016), consisting of 4016 X-ray emitters over a 2.2 square degree area; the catalog is the result of a 4.6 Ms program with a depth of $\simeq 160$ ks in the central 1.5 square degree and a $\simeq 80$ ks depth in the surrounding area. The limiting depth corresponds to a flux of 2.2×10^{-16} erg cm $^{-2}$ s $^{-1}$ in the soft (0.5 – 2 keV) X-ray energy band and to 1.5×10^{-15} erg cm $^{-2}$ s $^{-1}$ in hard (2 – 10 keV) X-rays, while the depth for the full (0.5 – 10 keV) X-ray band is 8.9×10^{-16} erg cm $^{-2}$ s $^{-1}$. The catalog provides a considerable volume of information, including optical counterparts for 3899 (97%) of the 4016 sources, a spectroscopic classification (BLAGN, non-BLAGN, star) for $\approx 42\%$ of the sample, and a photometric classification (unobscured AGN, obscured AGN, galaxy, star) based on SED fitting for 96% of the sample;
- the *XMM*-COSMOS Point-like Source catalog (Brusa et al., 2010), including 1674 X-ray sources with an optical counterpart. The corresponding program in this case is shallower ($\simeq 60$ ks) and a little less wide (2 square degree area) than the *Chandra*-COSMOS Legacy Survey; the flux limit in the 0.5 – 2 keV, 2 – 10 keV, and 5 – 10 keV energy bands is, respectively, $\approx 1.7 \times 10^{-15}$ erg cm $^{-2}$ s $^{-1}$, $\approx 9.3 \times 10^{-15}$ erg cm $^{-2}$ s $^{-1}$, and $\approx 1.3 \times 10^{-14}$ erg cm $^{-2}$ s $^{-1}$ over 90% of the area. Spectroscopic classification (BLAGN, NLAGN, normal/star-forming galaxy) is avail-

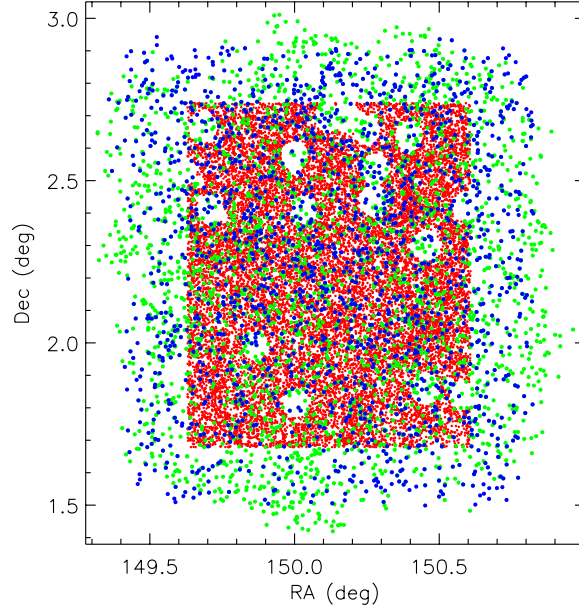


Figure 4.1: Overlap of the COSMOS sources from the VST main sample (small red dots) and from the *Chandra* (large green dots) and the XMM (large blue dots) catalogs. The two X-ray imaged fields roughly overlap, and they enclose the VST-COSMOS field. The voids in the VST catalog correspond to regions that were masked as affected by the presence of bright star halos (see Section 3.3).

able for about half the sample, and a best-fit SED template by Salvato et al. (2011) is provided for 97% of the sample.

In Figure 4.1 we report the overlap of the COSMOS field as imaged by the VST, the *Chandra*, and the XMM observatories.

The two X-ray catalogs provide information for 2371 X-ray sources in the VST-COSMOS FoV and with coordinates of their optical counterparts reported in the X-ray catalogs themselves. Nevertheless, not all of them have a VST counterpart; specifically, we found:

- 675 X-ray sources with magnitude $r(\text{AB}) < 23$ mag;

- 1471 out of 2371 sources with magnitude $r(\text{AB}) > 23 \text{ mag}$ ¹, hence not taken into account in what follows;
- 223 sources falling onto masked areas, hence with no counterpart in the VST-COSMOS main sample;
- one source with a VST counterpart but detected in less than six epochs and hence excluded from the main sample;
- one source detected in the stacked image, but too faint ($r(\text{AB}) > 24 \text{ mag}$) to be detected in any of the epochs.

Among the 675 X-ray sources with magnitude $r(\text{AB}) < 23 \text{ mag}$ and falling in the VST-COSMOS FoV, there are 12 for which we do not find a VST counterpart: this subsample includes four objects with no magnitude estimate in either of the two X-ray catalogs, and two of them do not have a counterpart in any of the other available COSMOS catalogs, while the other two only have a counterpart in the S-COSMOS Catalog (e.g., Sanders et al. 2007; see Section 4.3.2) so it is possible that they are spurious sources or sources that are very bright in the X-rays and very faint at all other wavelengths, and are hence beyond the depth of the VST; the remaining eight objects are located in the very proximity of a masked area or of a brighter source. As a consequence, the sample of X-ray sources that have a VST counterpart and fulfill our selection criteria consists of 663 objects (hereafter, X-ray sample); this means that 3% of the 19226 sources in the main sample of VST sources have an X-ray counterpart. We note that, although 12 of the objects in the X-ray sample are classified as stars, we chose not to exclude them from our analysis. The match of the secure sample with the X-ray sample revealed that

¹ $r(\text{AB})$ magnitudes from Capak et al. (2007) are reported in the XMM catalog and are available for 98% of the sources.

210 AGN candidates (79% of the secure sample: hereafter, X-ray AGN candidates) have an X-ray counterpart; this implies that the X-ray sources with an optically variable counterpart in our catalog are 210 out of 663 (32%); among them, only one source is photometrically classified as a star.

We point out that, while the mentioned 79% fraction is roughly the same as the one obtained from the *D0* data analysis (i.e., 76%), the fraction of X-ray emitters that are now classified as AGN candidates on the basis of our variability selection is almost three times the one that we got from *D0* data, which was 11%. This is something that we predicted in De Cicco et al. (2015) on the basis of the red noise power spectrum of AGNs, which allows to retrieve a larger number of variable sources if the baseline is longer. We will go back to this point in the final discussion (Chapter 6). It is also worth noting that only one of the sources in the temporary sample of AGN candidates and that we excluded from our secure sample has a counterpart in the X-ray sample.

In what follows, unless otherwise stated, we use Chandra values preferentially as the catalog is deeper and larger than the XMM catalog and, if data are not available, we resort to the XMM catalog. All of the 210 X-ray AGN candidates have a counterpart in the Chandra catalog.

The Chandra catalog reports X-ray luminosities L_x , together with both spectroscopic and photometric redshifts, for most of the sources and, in particular, for 208 of the X-ray AGN candidates: for all of them, the hard L_x is $> 10^{42}$ erg s $^{-1}$, which is typical of AGNs, while non-active galaxies are characterized by $L_x < 10^{42}$ erg s $^{-1}$ (e.g., Brandt & Hasinger, 2005). The plot of the hard X-ray luminosity as a function of redshift is reported in Figure 4.2, and it clearly corroborates the reliability of our candidate selection.

The ratio of the X-ray to optical flux (X/O) of a source is defined as

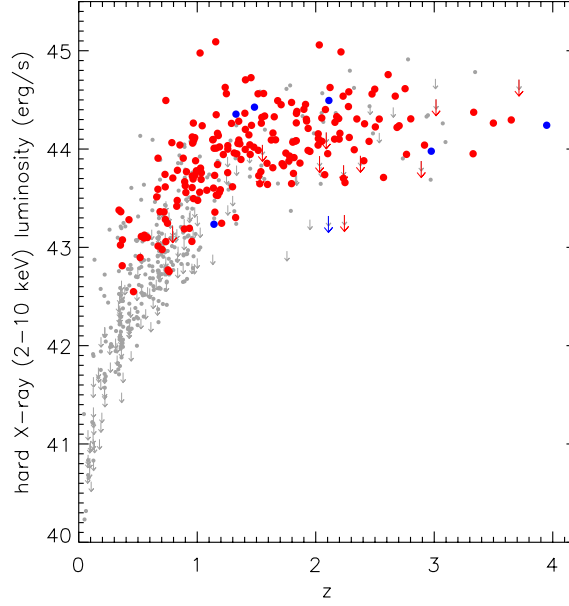


Figure 4.2: Hard (2 – 10 keV) X-ray luminosity L_x vs. redshift z for the 210 X-ray AGN candidates (larger symbols), compared to the overall X-ray population. For all but seven sources, spectroscopic redshifts are available in the *Chandra* catalog, and such objects are represented by red symbols; blue symbols stand for those sources for which only photometric redshifts are available, while downward arrows indicate that only upper limits of the flux values are available. Smaller grey symbols in the background represent the X-ray sample, reported as a reference population.

(Maccacaro et al., 1988)

$$X/O = \log(f_x/f_{opt}) = \log f_x + \frac{\text{mag}_{opt}}{2.5} + C, \quad (4.1)$$

f_x being the X-ray flux measured in the chosen energy band, while mag_{opt} is the optical magnitude at the chosen wavelength, and C is a constant which depends on the magnitude system adopted for the observations. While for stars and non-active galaxies the X/O is typically < -2 (e.g., Mainieri et al., 2002; Xue et al., 2011), AGNs are usually characterized by $-1 \leq X/O \leq 1$, hence we can use the X/O of our sources with an X-ray counterpart to

investigate their nature. In Figure 4.3 we show the hard ($2 - 10$) X-ray flux *vs* r -band magnitude for the sources in the secure sample that have an X-ray counterpart. It is apparent that all but two of the X-ray AGN candidates place themselves within the AGN locus. This means that, if no additional information about our secure sample were available, on the sole basis of the X/O diagram we could be confident that 78% (208/265) of the candidates are indeed AGNs. We point out that there is always some degree of uncertainty in the definition of the X/O of a source, because of the intrinsic variability of the source itself combined to the non-simultaneity of the X-ray and optical observations.

4.2 Spectroscopic and Photometric Classification

We already mentioned (Section 4.1) that both *Chandra* and XMM catalogs provide a spectroscopic and a photometric classification for part of the X-ray emitters in each of the catalogs (Marchesi et al., 2016). Objects with a spectroscopic classification are labeled as BLAGNs if their spectra show at least one broad ($\text{FWHM} > 2000 \text{ km s}^{-1}$) emission line, while non-BLAGNs could be NLAGNs as well as star forming galaxies: this is because most of the sources at issue are characterized by low S/N spectra or because the waveband in which the spectra are obtained does not allow to make use of optical emission line diagnostics which would help separating the two classes of objects. The photometric classification is derived through best-fit templates of the broadband SEDs, and sources are divided into unobscured AGNs, obscured AGNs, non-active galaxies, and stars. A cross-match of the two classifications revealed that 82% of the BLAGNs correspond to unobscured

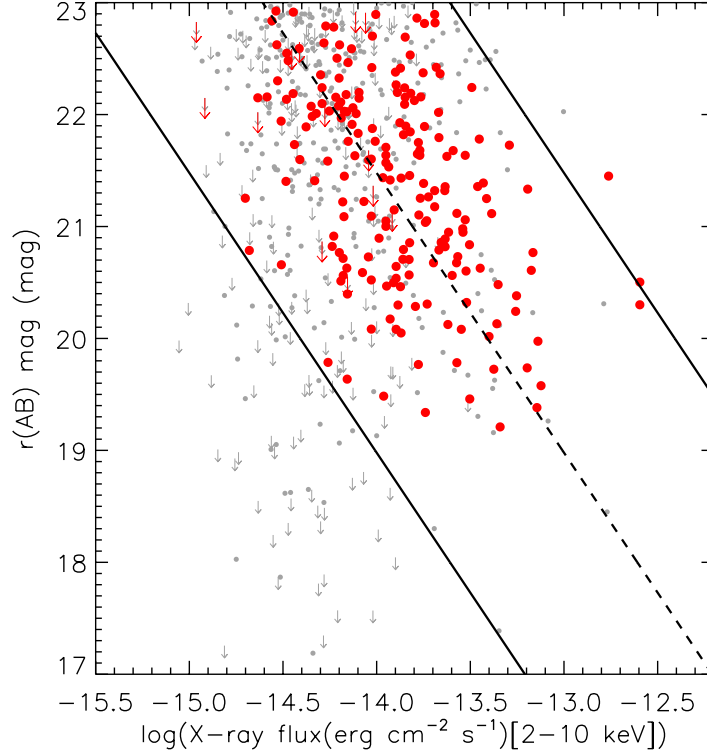


Figure 4.3: $r(\text{AB})$ magnitude vs. hard (2-10 keV) X-ray flux for the 210 X-ray AGN candidates (larger red symbols). The smaller grey symbols in the background are shown as a reference population and stand for the X-ray objects in the X-ray sample. Downward arrows indicate that only upper limits of the X-ray flux values are available. The dashed line corresponds to $X/O = 0$; the lower and upper solid lines represent $X/O = -1$ and $X/O = 1$, respectively, and define the AGN locus.

AGNs; the match is not higher because BLAGN SEDs, especially in case of low-luminosity AGNs, suffers from stellar light contamination; non-BLAGNs are matched to obscured AGNs in 23% of the cases and to galaxies in 74% of the cases.

The XMM catalog classifies sources as BLAGNs, NLAGNs, and non-active galaxies (Brusa et al., 2010). BLAGNs must fulfill the same criterion as in the *Chandra* catalog; NLAGNs typically have spectra characterized by

unresolved high-ionization emission lines with line ratios suggesting AGN activity, or have rest-frame hard X-ray luminosities $L_X > 2 \times 10^{42} \text{ erg s}^{-1}$, while non-active galaxy spectra are generally consistent with those of star-forming or normal galaxies, and have rest-frame hard $L_X < 2 \times 10^{42} \text{ erg s}^{-1}$, when detected in the hard X-rays; part of the best-fit SED templates correspond to Type 1 and Type 2 AGNs.

In our X-ray sample (which –we remind– consists of all the VST sources with an X-ray counterpart, regardless their optical variability) there are 654 out of 663 sources with spectroscopic and/or photometric classification from either of the two catalogs; among these, we have 394 confirmed AGNs which, specifically, are split into 247 Type 1 AGNs and 147 Type 2 AGNs. In order to assign our sources to one or the other class, we always use the *Chandra* catalog as a reference, and we assume that spectroscopic classification is more reliable than photometric classification. While the definitions for Type 1 AGNs are the same in the two X-ray catalogs, the label “non-BLAGN” in the *Chandra* catalog is ambiguous, and hence not sufficient to classify a source.

Our X-ray sample of 210 AGN candidates turns out to consist of 181 sources classified as Type 1 AGNs, 17 sources classified as Type 2 AGNs, one star, four non-active galaxies, and seven sources for which classifications is uncertain. Nonetheless, we note that the 11 sources not classified as Type 1/2 AGNs have an X-ray luminosity $L_X > 10^{42} \text{ erg s}^{-1}$, and lie onto the AGN stripe in the X/O diagram; moreover, half of them have a spectroscopic classification (which is non-BLAGN, from the *Chandra* catalog, for all of them and, for four of them, there is additional spectroscopic classification as NLAGNs from the XMM catalog), hence we do not exclude them from our secure sample.

Defining the completeness as the ratio of confirmed AGN candidates to expected AGNs, the above mentioned numbers allow us to compute the completeness of Type 1 and Type 2 AGNs, which are $181/247 = 73\%$ and $17/147 = 12\%$, respectively. These fractions are larger than those obtained in De Cicco et al. (2015), which were 26% and 6%, respectively. This is another effect of extending the time baseline (see Chapter 6).

4.3 Color-based Classification

Diagrams comparing two different source colors are widely used with the purpose of disentangling different classes of objects, based on the fact that sources tend to occupy distinct loci on such diagrams, depending on their nature.

4.3.1 Optical-IR Diagnostic

Nakos et al. (2009) propose the use of a diagram where the two colors are $r-z$ and $z-k$: such a choice is suitable for QSO identification, as their SEDs are characterized by an excess of emission in the k band which reflects into a redder $z-k$ color if compared to stars, and a redder $r-z$ color if compared to galaxies. On an $r-z$ vs. $z-k$ plane, stars define a rather sharp sequence, while galaxies tend to place themselves on a more scattered region.

In order to obtain such a diagram, we made use of data from two additional catalogs of the COSMOS field:

- the COSMOS ACS catalog (Koekemoer et al., 2007; Scoville et al., 2007) from the *Hubble* Space Telescope, obtained from 575 pointings of the Advanced Camera for Surveys (ACS). The catalog includes a stellarity index (see Section 3.1), which will prove useful in the following;

- the COSMOS Intermediate and Broad Band Photometry catalog, which provides the magnitudes we need in the r , z , and k bands, down to a magnitude $r(\text{AB}) \approx 29$ mag. The catalog includes photometry in 25 optical/NIR narrow, intermediate, and broadband filters from u to K_s , and supersedes the one by Capak et al. (2007), with improved source detection and photometry extracted in $3''$ -diameter apertures.

Our r - z vs. z - k diagram is shown in the upper panel of Figure 4.4. The availability of a morphology indicator allows to identify different classes of objects on the diagram. In addition to the stellar sequence and the galaxy region, a third population of sources can be located below the galaxy area: it consists of QSO-like objects, i.e., compact sources –as suggested by their stellarity index– with non-stellar colors, typical of extended objects –as indicated by their position on the diagram–, suggesting that we are indeed facing galaxies. The chosen diagram allows to classify as QSO-like AGNs 173 out of 265 (65%) of the AGN candidates in the secure sample; this number does not include sources with $z - k < 0.05$, after the star/AGN separation criteria discussed by Nakos et al. (2009), converted to our AB magnitude system. In the lower left panel of Figure 4.4 we show the same plot as in the upper panel, but we add crosses to mark sources that were already confirmed to be AGNs on the basis of previously discussed diagnostics. In the lower right panel the diagram obtained from the analysis of $D0$ data is shown: in this case as well, the fraction of AGNs confirmed on the basis of the r - z vs. z - k diagram is 65%.

Since for our analysis we could take advantage of a large number of diagnostics to confirm our candidates, only seven sources turned out to be confirmed *ex novo* as AGNs; nevertheless we point out that, if no additional information about X-ray emission, nor spectra, were available for our sources,

the color vs. color diagram here discussed would allow us to confirm 65% of our secure sample as AGNs on the basis of their variability, color, and compactness.

Although NIR data allow to separate AGNs from galaxies better than traditionally used optical colors (e.g., $U-B$ vs. $B-V$ diagrams), there could be a significant fraction of AGNs –including fainter ones, where the host galaxy contamination is more severe– whose colors are consistent with those of non-active galaxies and that hence do not lie in the QSO region in the proposed diagram and cannot be identified by means of the used colors.

4.3.2 Mid Infrared Diagnostic

Lacy et al. (2004) discuss a color vs. color diagram which makes use of mid infrared (MIR) colors from the Infrared Array Camera (IRAC; e.g., Fazio et al. 2004) of the *Spitzer* Space Telescope to select AGNs. IRAC operates simultaneously at four different wavelengths, namely 3.6, 4.5, 5.8, and 8.0 μm ; the strength of the proposed diagnostic lies in the fact that the 8.0 μm /4.5 μm and the 5.8 μm /3.6 μm ratios allow, when combined in a diagram, to separate sources with different continuum emissions. If we take as a reference the diagram obtained from our analysis, shown in the upper panel of Figure 4.6, several distinct loci are visible: stars, as well as low-redshift galaxies, are characterized by bluer colors on both axes, and define the blob that can be observed in the lower left part of the plot; galaxies having $z \simeq 0.2$ are preferentially found on the roughly vertical sequence corresponding to blue colors in 5.8 μm /3.6 μm and red in 8.0 μm /4.5 μm if the bulk of their MIR emission is redshifted in correspondence of the 8.0 μm filter; a third sequence is defined by QSOs, characterized by red colors on both axes, unless there is a significant stellar contribution to the emission,

which would shift the ascissa color towards bluer values. By defining color criteria, Lacy et al. (2004) delineated boundaries to the region where QSOs generally place themselves on the diagram at issue; such boundaries were slightly modified in a subsequent work (Lacy et al., 2007), which is the one we will refer to in the following (hereafter, Lacy region).

A more substantial revision of the IRAC selection criteria is presented in Donley et al. (2012), aiming at reducing the heavy contamination by star-forming galaxies, which is particularly significant at high redshift ($z \gtrsim 2$). We will label the corresponding AGN locus “Donley region”. In Figure 4.5, from Donley et al. (2012), IRAC colors of composite SEDs constructed with different AGN contributions are shown in the diagram discussed above (Lacy et al., 2004), in the redshift range $0 - 3$; the Lacy region is also shown. We notice that, as the AGN contribution becomes dominant, the IRAC colors of the corresponding sources tend to shift inside the Lacy region and towards bluer colors, and tend to get closer to the IRAC power-law locus, i.e., the line where an ideal source with perfect IRAC power-law SED would fall. However, the position of a source on such a diagram depends not only on the relevance of its AGN component, but also on redshift, reddening, and host galaxy type.

In order to reproduce the diagnostic by Lacy et al. (2007) discussed above, we made use of the S-COSMOS Catalog (e.g., Sanders et al., 2007), providing information about more than 345000 sources imaged over the 2 square degree area of the Cosmic Evolution Survey, and including photometry in the four IRAC channels. Our IRAC diagram in each the three panels of Figure 4.6 includes both boundaries from Lacy et al. (2007) and Donley et al. (2012) in order to allow a more accurate analysis. In particular, the diagram shown in the upper panel refers to $D1$ analysis. IRAC fluxes are available for all but

nine of the 265 sources in our secure sample; we find that 72% (190/265) of the sample lies within the Lacy region, and this percentage includes three sources confirmed *ex novo*, while 34% (90/265) of the secure sample lies in the Donley region, and in this case for all these sources we have additional confirmation from other diagnostics. In the lower left panel of Figure 4.6 we show the same plot as in the upper panel, but include crosses to mark sources that were already confirmed to be AGNs on the basis of previously discussed diagnostics. In the lower right panel the same diagram obtained from the analysis of *D0* data is shown: in this case the fraction of AGNs candidates lying within the region defined by Lacy et al. (2007) is 82%, while 60% of them lies in the region defined by Donley et al. (2012).

From the diagram shown in the lower left panel it is apparent that a relevant fraction of AGNs confirmed by means of other diagnostics fall outside the regions defined by Donley et al. (2012) and/or Lacy et al. (2007): this happens when the host galaxy emission is substantial and dims the light contribution from the active nucleus, and hence AGNs are hidden by colors typical of non-active galaxies and cannot be identified by means of the chosen colors.

4.4 The Structure Function of Confirmed AGNs

In Section 1.3 we mentioned that the time dependence of the variability of a sample of sources can be parametrized by means of the structure function. Several definitions for it are proposed in the literature (e.g., Simonetti et al., 1985; Di Clemente et al., 1996; Graham et al., 2014); basically, given a sample of sources, the SF shows the ensemble rms magnitude difference as a

function of the time lag between different epochs, hence we shall define it as

$$SF = \sqrt{\langle [\text{mag}(t+\Delta t) - \text{mag}(t)]^2 \rangle} ; \quad (4.2)$$

$\text{mag}(t)$ and $\text{mag}(t+\Delta t)$ are two observations of the same source at different times, and the time difference between the two is $\Delta t \leq \tau$, where τ is the time lag over which we perform the average. The SF allows to deal with large samples of objects and characterize their overall variability; its strength lies in the fact that the different time lags over which we measure variability are mutually independent, while this does not hold if we take each source into account individually (e.g., de Vries et al., 2005).

Our analysis is limited to the sample of VST-COSMOS sources that are spectroscopically confirmed AGNs, regardless their optical variability. All the selected sources have $r(\text{AB}) \leq 23$ mag; the confirmation of their nature is based on the spectral classification reported in the *Chandra*-COSMOS Legacy Catalog (Marchesi et al., 2016; Civano et al., 2016) or the XMM-COSMOS Point-like Source catalog (Brusa et al., 2010), which we introduced in Section 4.1. We selected all the objects classified as broad-line AGNs in the *Chandra* catalog or as broad-/narrow-line AGNs in the XMM catalog, on the basis of the discussion in Section 4.2 about the spectral classification in the two X-ray catalogs; the selected sample consists of 394 sources.

In Figure 4.7 we show the SF obtained for our sample of AGNs. Time differences are rest frame, and were obtained by dividing the time corresponding to each epoch by $1+z$, where z is the redshift of the corresponding source. Three distinct regions can be identified on the plot: when time lags are too short to detect variability, all we can see is a plateau corresponding to $\sqrt{2}\sigma_{\text{noise}}$, σ_{noise} being the rms of the measurement noise of the instrument (e.g., Hughes et al., 1992); from time lags of some days on, the SF is approximately linear, with a 0.40 ± 0.08 slope; the final drop is not real, but it is

an effect of the poor sampling at large time lags. The linear trend originates from the fact that the difference between two measured magnitudes of an AGN increases with the time difference between the two corresponding observations. The measured slope is fairly consistent with the values reported in Schmidt et al. (2010); Bauer et al. (2009); Vanden Berk et al. (2004), i.e., 0.45 ± 0.02 , 0.3607 ± 0.0075 , and 0.336 ± 0.033 , respectively.

An estimate of the noise contribution can be obtained by means of the SF of a sample of non-variable sources. The sample that we selected consists of VST-COSMOS objects which are non-variable optically and not classified as AGNs in the above mentioned X-ray catalogs; an estimate of their spectroscopic redshift is available from at least one of the X-ray catalogs or from the z-COSMOS catalog (Lilly et al., 2007), and we required the significance of the sources (see Section 3.4) to be < 2 . A resampling on the basis of the magnitude of the sources in the AGN sample was performed, so that both sets of objects have the same magnitude distribution. The SF obtained for this sample of objects is shown in Figure 4.7 as well. In the case of non-variable sources we expect an approximately flat SF, where the difference between the two magnitudes in each pair of compared observations is due to the measurement noise. Since the observations of both AGNs and non-variable objects come from the same data, we expect the two structure functions to overlap at short time lags, and from Figure 4.7 we can indeed see a quite good overlapping in the left region of the plot.

The existence of correlations between the AGN structure function and some properties of the central SMBH, such as its mass or accretion rate (expressed by the Eddington ratio E_r), has been widely investigated in the last decades, especially in the X-rays, which map the AGN inner regions (e.g., McHardy et al., 2006, and references therein). An anti-correlation

between variability and luminosity is generally observed, and it is usually considered to be an effect of the anti-correlation of variability with the BH mass, while conflicting results are found about a possible anti-correlation with the accretion rate. In Simm et al. (2016) the search for such correlations concerns AGNs in the XMM-COSMOS catalog from Brusa et al. (2010), and AGN variability –quantified by the excess variance– is found to anti-correlate with the Eddington ratio, while no dependence on black hole mass is observed.

In order to unearth possible correlations in our sample of AGNs, we selected those, among the 394 spectroscopically confirmed ones, for which an estimate of the BH mass and the Eddington ratio² are available in the catalogs presented in Lusso et al. (2012) and Lusso et al. (2010), concerning a sample of 929 Type 1 and Type 2 AGNs and a sample of 545 Type 1 AGNs (partly overlapping with the first sample), respectively, both from the XMM-COSMOS survey.

As a further step, we divided the subsample of AGNs with mass and E_r estimates into three additional subsets of AGNs: BH masses range $10^{7.66}$ to $10^{9.26} M_{\odot}$; the three subsets (hereafter, BH mass subsamples) were obtained dividing into equal parts the whole BH mass range. We did the same with Eddington ratios, which range $10^{-1.99}$ to $10^{-0.05}$, thus obtaining three subsamples (hereafter, E_r subsamples) independent on the ones defined on the basis of BH masses. We then computed, for each set of subsamples, the structure function in the three different bins; our preliminary results are shown in Figure 4.8. We point out that this part of the work is still ongoing and needs some refinement. While no evidence for a correlation between AGN variabil-

²The Eddington ratio is defined as $E_r = L_{\text{bol}}/L_{\text{Edd}}$, where L_{bol} is the already mentioned bolometric luminosity and L_{Edd} is the Eddington luminosity.

ity and BH mass emerges from the corresponding diagram (top panel), an anti-correlation with the Eddington ratio is apparent in the region where the SF is linear (bottom panel); this is consistent with the findings from Simm et al. (2016).

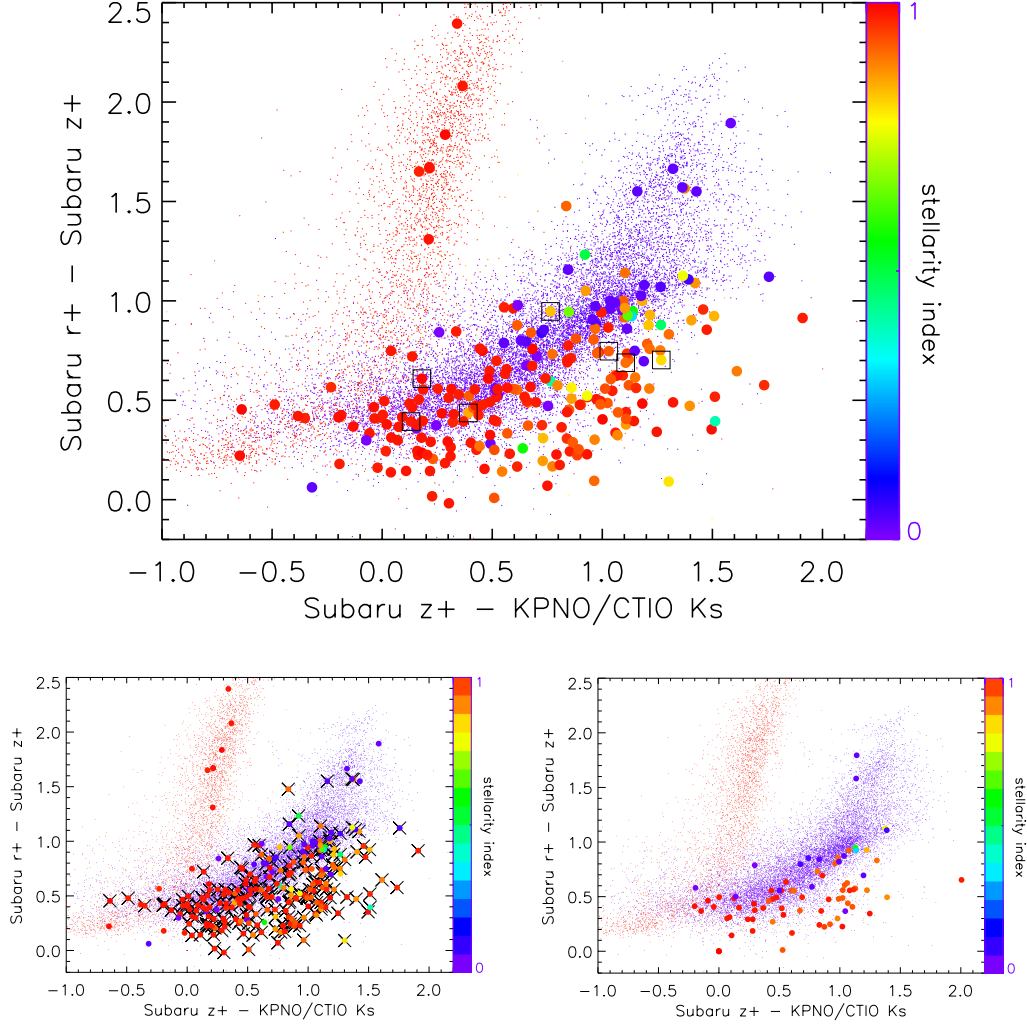


Figure 4.4: *Upper panel:* r - z vs. z - k diagram for the 260 out of 265 AGN candidates in the secure sample (larger dots) for which r , z , k magnitudes and stellarity index are available. The smaller dots in the background represent all the objects detected in the VST-COSMOS field for which stellarity index and color information are available, and are shown as a reference population. Both large and small dots are color-coded according to their stellarity index (right vertical axis). It is apparent that the sources in the plot define two distinct loci: one for stars (red background dots), and one for galaxies (violet background dots); moreover, a third class of objects, i.e., QSO-like AGNs, is identified: the corresponding sources lie onto and below the galaxy locus but have stellarity indices typical of compact sources (≥ 0.8). Boxes indicate the seven AGN candidates confirmed *ex novo* by means of such diagram. K_s magnitudes are by McCracken et al. (2010). *Lower panels:* same plot as in the upper panel, with crosses indicating the sources in the secure sample that are already confirmed to be AGNs on the basis of previously discussed diagnostics (*left*); same plot as in the upper panel, obtained from the analysis of $D0$ data (*right*).

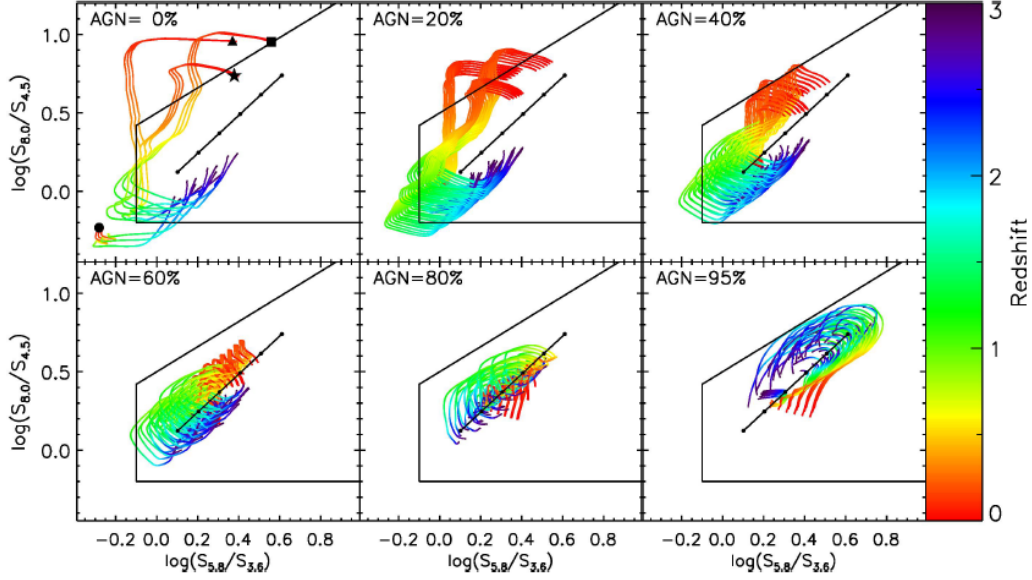


Figure 4.5: IRAC colors of AGN/galaxy composite SEDs, in the redshift range 0 – 3. Templates were used to represent a star-forming galaxy (square), a starburst galaxy (star), a normal star-forming spiral galaxy (triangle), and an elliptical galaxy (circle; see Donley et al. 2012 for further details). Large symbols in the upper left panel mark the corresponding family of purely star-forming templates at redshift $z = 0$; in each of the remaining panels, a different fraction of AGN contribution, defined in the range $1 - 10\mu\text{m}$, is taken into account. The region defined by the black solid line is the already mentioned Lacy region. The black line inside such region represents the IRAC power-law locus from $\alpha = 0.5$ (lower left) to $\alpha = 3.0$ (upper right; see Equation 1.1). Templates of purely star-forming galaxies do not cross the power-law locus, but they fall inside the AGN region at both low and high redshift. As the AGN contribution becomes more relevant, the SEDs move inwards and redwards toward the power-law locus. Credit: Donley et al. (2012).

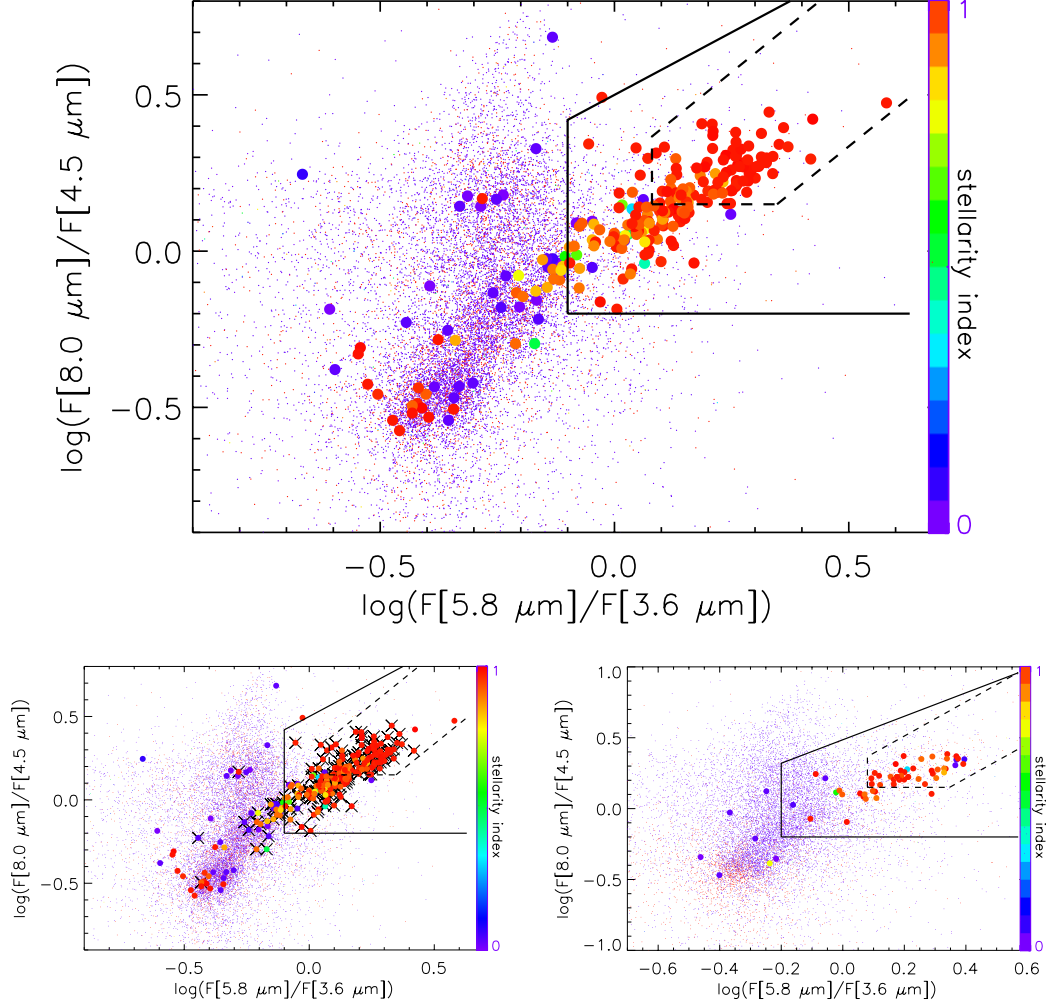


Figure 4.6: *Upper panel*: MIR diagram where colors are obtained as ratios of the fluxes in the four IRAC channels. Larger dots represent the AGN candidates in our secure sample, and are color-coded on the basis of their stellarity index in the COSMOS ACS catalog (see Section 4.3.1). Smaller dots in the background are the VST-COSMOS sources with an IR counterpart in the S-COSMOS catalog, and are reported as a reference. The solid line delineates the region where AGNs are typically found according to Lacy et al. (2007), while the dashed line defines the less contaminated AGN region identified in Donley et al. (2012). A blob characterized by a high concentration of stars can be seen in the lower left part of the diagram, while non-active galaxies tend to occupy the roughly vertical, scattered sequence defined by small violet dots. *Lower panels*: same plot as in the upper panel, with crosses indicating the sources in the secure sample that are already confirmed to be AGNs on the basis of previously discussed diagnostics (*left*); same plot as in the upper panel, obtained from the analysis of *D0* data (*right*).

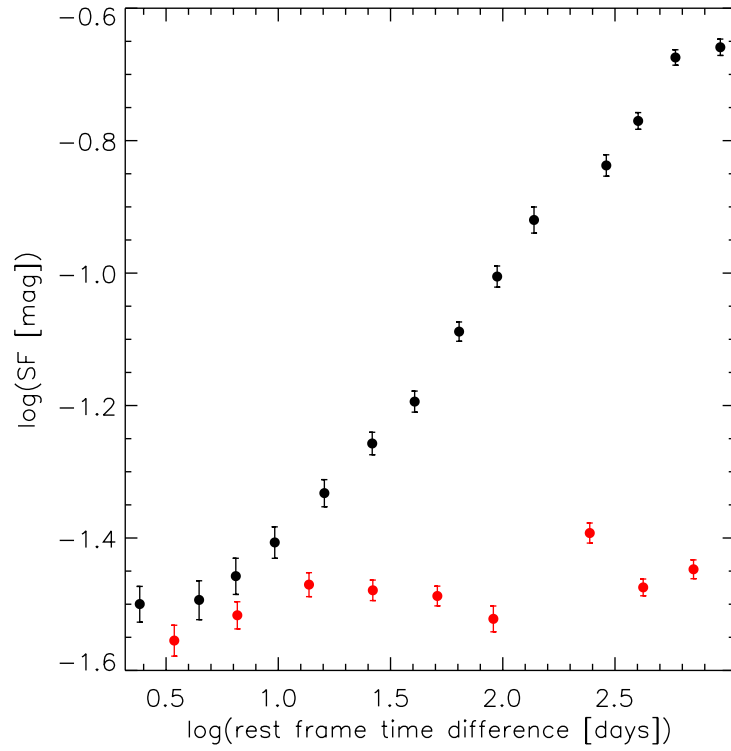


Figure 4.7: Structure function as defined by Equation 4.2 for our sample of AGNs (black dots). The SF of our sample of non-variable sources is also shown (red dots). The two SFs fairly overlap in the region corresponding to short time lags.

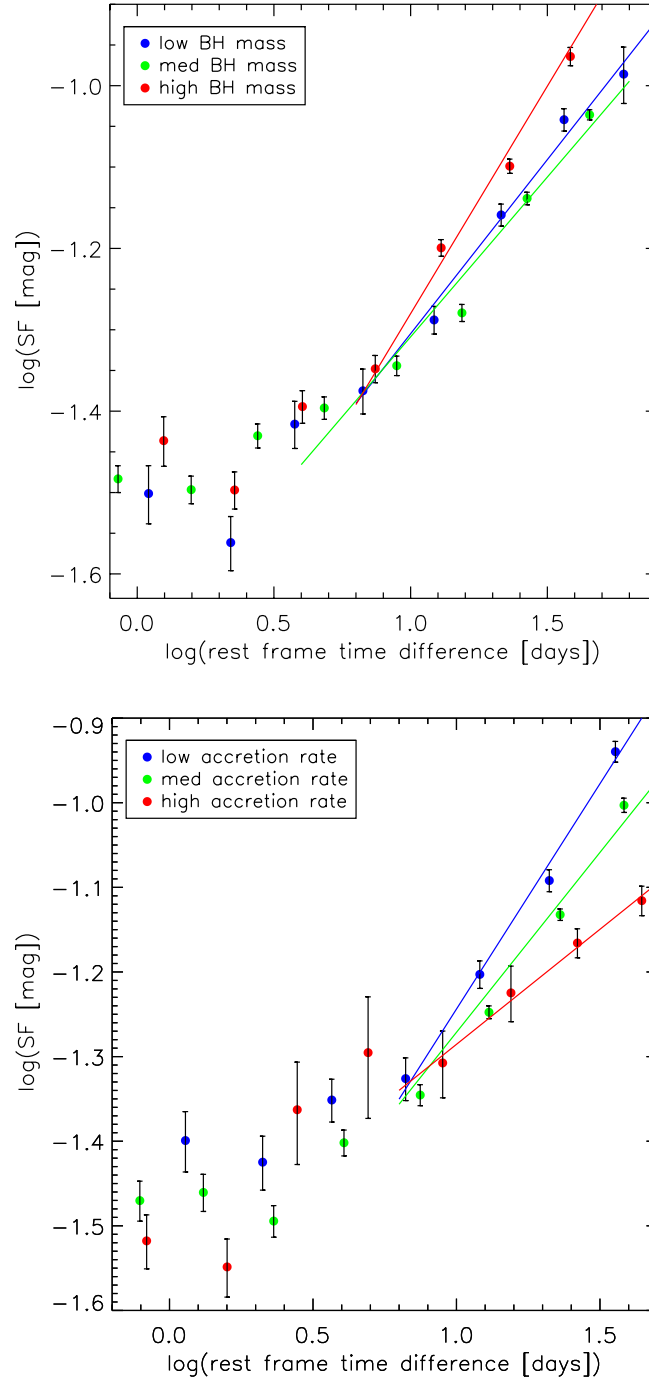


Figure 4.8: Structure function for the BH mass subsamples in three different mass bins (*top panel*) and for the E_r subsamples in three different accretion rate bins (*bottom panel*). Blue corresponds to the lowest bin, green to the middle one, and red to the highest bin in either plot.

Chapter 5

Extension of the Variability Analysis to the CDFS

In Chapter 3 we anticipated that the analysis described so far was also performed over the CDFS data obtained from the VST-SUDARE survey, in order to strengthen the results obtained from the COSMOS data analysis and make use of additional selection techniques to be compared to the AGN selection based on optical variability. A first analysis concerned a 2 square degree area, and results were presented in Falocco et al. (2015); a subsequent analysis over two additional square degrees is ongoing and will be described in Poulain et al. (in prep.). In what follows we briefly introduce the dataset and present the main findings from Falocco et al. (2015).

In Figure 5.1 we show the CDFS area imaged by the VST, together with the areas covered by overlapping surveys. X-ray coverage in this case is limited to the Extended Chandra Deep Field South (ECDFS; e.g., Lehmer et al. 2005), i.e., a region which is roughly 1/8 of the area analyzed in Falocco et al. (2015). X-ray data are from the *Chandra* survey of the 4 Ms CDFS observations (Xue et al., 2011; Rangel et al., 2013) and from the 250 ks CDFS

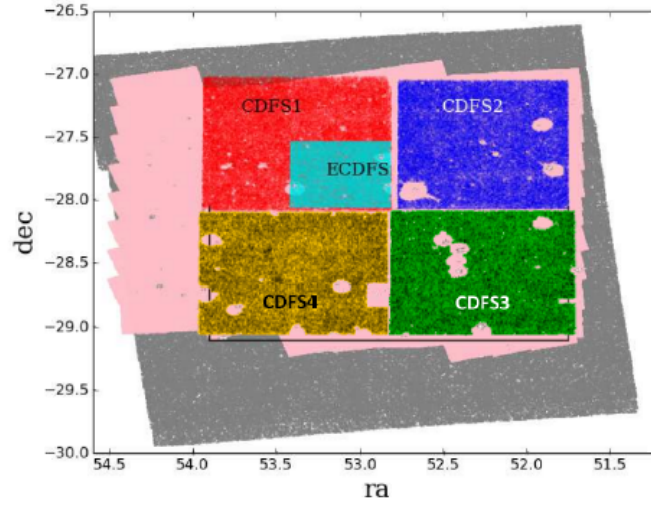


Figure 5.1: The Chandra Deep Field South as imaged by the VST (red, blue, green, and yellow are for VST-CDFS1, VST-CDFS2, VST-CDFS3, VST-CDFS4, respectively), SWIRE (grey), SERVS (pink), and ECDFS (cyan). VST-CDFS3 and VST-CDFS4 areas are being currently investigated, and results are being gathered in Poulain et al. (in prep.). As well as in Figure 4.1, the voids in the VST-imaged areas represent regions that we masked, mostly because of the presence of bright star halos. Celestial coordinates are expressed in degrees. Credit: Poulain et al. (in prep.), adapted from Falocco et al. (2015).

observations (Lehmer et al., 2005; Virani et al., 2006). For ECDFS sources, SED coverage is also available; both X-ray and SED data were collected in Hsu et al. (2014), and we made use of the information they provide. While X-ray coverage is limited with respect to the COSMOS field, there is full overlapping of the VST region with those imaged by IR/optical surveys, namely the *Spitzer* Wide-Area InfraRed Extragalactic (SWIRE) survey (Lonsdale et al., 2004) and the *Spitzer* Extragalactic Representative Volume Survey (SERVS; Mauduit et al. 2012); this allowed us to compare the strength of AGN selection on the basis of optical variability to additional selection techniques based on the use of IR data.

The analysis was performed on r band data for the same reasons discussed

when describing the VST-COSMOS dataset (Section 2.2); observations, as well as COSMOS field ones, were generally obtained every three days, depending on a number of constraints. The analyzed dataset consists of 27 observations spanning five months for the CDFS1 and 22 observations over four months for the CDFS2. The data reduction and sample selection procedures are essentially the same as the ones described in Chapter 3.

5.1 Main Results

The selection method based on optical variability returned a sample of 175 AGN candidates over the two surveyed areas, once the corresponding samples were cleaned from potentially spurious candidates. For 158 of them we found an IR counterpart in the SERVS and/or SWIRE catalogs¹; three of the remaining 17 sources fall out of the surveyed area, while seven objects turned out to be SNe, and it is very likely that they are hosted by normal galaxies with no significant IR emission; this could explain the lack of a counterpart for them in the IR catalogs. The sources with an IR counterpart and for which photometry at 3.6, 4.5, 5.8, and 8.0 μm is available are 115 out of 158. Fifteen out of 175 AGN candidates fall in the ECDFS area. Among them, 12 sources have an X-ray counterpart. Here we discuss in short the main results obtained:

- the visual inspection of the AGN candidate light curves, together with the cross-correlation with the findings from the SN search and analysis over the VST-CDFS (Cappellaro et al., 2015), allows to estimate a

¹Since the two catalogs overlap over an area of 6 square degrees, we built a new catalog from their match in order to store together all the available information about the sources in that area, and we shall hereafter refer to it as the SERVS+SWIRE catalog.

- 12% (21/175) contamination from SNe; such number includes the three sources in the ECDFS with no X-ray counterpart;
- the contamination from stars is 3% (6/175); this was derived by means of an optical/IR diagnostic proposed in Rowan-Robinson et al. (2005), where two colors, namely $r - i$ and the $3.6 \mu\text{m}$ to r band flux ratio, are compared on a diagram. Such a choice allows to separate stars from galaxies as the first ones are typically characterized by stronger IR emission. The diagram obtained for our sources is shown in Figure 5.2: it is apparent that two distinct classes of objects exist; the black line sets a more specific boundary between the two, in order to reduce mutual contamination; galaxies are located above and stars below the line. The required fluxes are available for 57 out of the 175 AGN candidates. Apart from the six sources classified as stars, there are 51 objects lying in the galaxy region, and they are mostly characterized by non-stellar colors; conversely, their stellarity index reveals they are mostly point-like in shape: both features are typical of AGNs;
 - the 12 sources with an X-ray counterpart (7% of 175) are classified as AGNs on the basis of their X-ray luminosity and X-ray to optical flux ratio: the X-ray luminosity is always $L_x > 10^{42} \text{ erg s}^{-1}$, and the X-ray to optical flux ratio of each of the 12 sources lies in the range $-1 \leq X/O \leq 1$; both features are typical of AGNs (see Section 4.1); furthermore, the SEDs of the 12 objects show a substantial contribution from AGN emission;
 - 58% (102/175) of the sample of AGN candidates lies in the region where, after Lacy et al. (2004, 2007), AGNs are typically found on the MIR diagram discussed in Section 4.3.2, comparing $8.0 \mu\text{m}/4.5 \mu\text{m}$ and

$5.8\ \mu m/3.6\ \mu m$ flux ratios. The diagram for CDFS data is shown in Figure 5.3. The stellarity index, combined with the two colors compared on the plot, shows that most of the compact AGN candidates place themselves in the region defined in Donley et al. (2012), while more extended sources can be found outside that region, i.e., where the contribution from starburst/host galaxy is expected to affect the source colors. Twelve sources lie outside the region defined by Lacy et al. (2004): five of them belong to the set of 12 confirmed SNe, while the lack of any specific AGN/star/SN signature does not allow to classify the remaining seven objects. The 102 sources confirmed as AGNs by means of the MIR diagnostic here discussed include the 12 objects with an X-ray counterpart we mentioned above;

- for 38 sources, including 12 of the confirmed SNe, no diagnostics other than light curves are available; the rest of the sample consists of objects for which at least one diagnostic is available, but they could not be confirmed by any of them.

Considerations about the results obtained from either work will be presented in Chapter 6, where we try and sketch the big picture of the method as a whole.

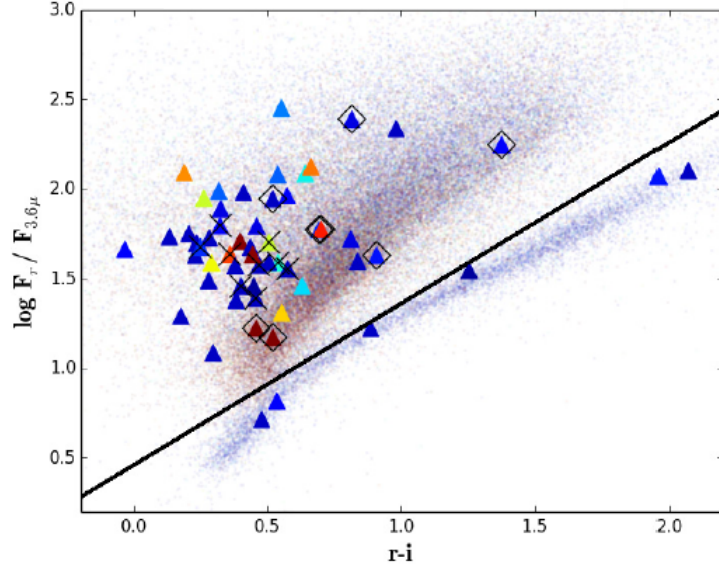


Figure 5.2: Optical/IR diagnostic for the 57 sources in the sample of 175 AGN candidates for which r , i , and $3.6\,\mu\text{m}$ fluxes are available from SERVS and/or SWIRE catalogs. The black line separates galaxies (above) from stars (below). Large triangles represent our AGN candidates; in particular, crosses mark those candidates having an X-ray counterpart, while diamonds indicate confirmed SNe. Smaller dots represent a reference population and consist of all the SERVES+SWIRE sources for which the required photometry is available. Objects are color-coded on the basis of their stellarity index (see Section 3.1) reported in the VST-CDFS catalogs; due to S. Falocco’s choice, here the color code is inverted with respect to plots concerning COSMOS data; as a consequence, the index grows going from red (indicating extended sources) to blue (corresponding to point-like objects). Credit: Falocco et al. (2015).

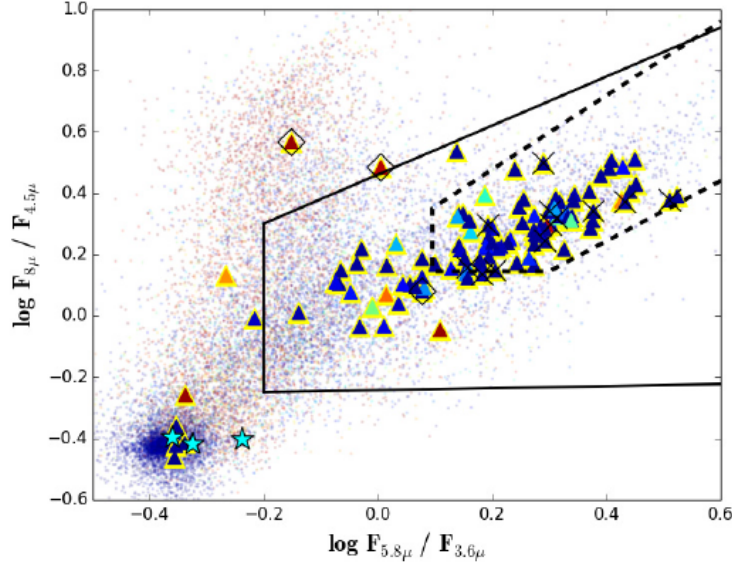


Figure 5.3: MIR diagram for the CDFS sources, as the one shown in Figure 4.6 for the COSMOS field. Colors are obtained as ratios of the fluxes in the four IRAC channels. Larger symbols represent the AGN candidates for which IR photometry is available; in particular, stars represent stars and triangles represent the rest of the sample of AGN candidates; among them, diamonds stand for SNe, and crosses mark candidates having an X-ray counterpart. Objects are color-coded on the basis of their stellarity index from the VST-CDFS catalog; once again, due to S. Falocco’s choice, here the color code is inverted with respect to plots concerning COSMOS data; as a consequence, the index grows going from red (indicating extended sources) to blue (corresponding to point-like objects). Smaller dots in the background are the sources in the SERVS+SWIRE catalog, and are reported as a reference. The solid line delineates the region where AGNs are typically found according to Lacy et al. (2004) (the region boundaries are a little different from the ones defined in Lacy et al. 2007, which we showed in Figure 4.6), while the dashed line defines the less contaminated AGN region identified in Donley et al. (2012). A blob characterized by a high concentration of stars can be seen in the lower left part of the diagram, while non-active galaxies tend to occupy the roughly vertical, scattered sequence defined by small reddish dots. Credit: Falocco et al. (2015).

Chapter 6

Discussion

In the first part of the thesis we selected a sample of 265 optically variable AGN candidates in the VST-COSMOS field, and made use of a number of multi-wavelength diagnostics to confirm their nature. In Table 6.1 we report the number of sources (including stars) confirmed by each diagnostic or by a combination of them.

We confirm as AGNs 220 sources (83% of 265) in the secure sample, thus proving the effectiveness of optical variability as an AGN selection method. This 83% represents the purity of our sample, where the purity is defined as the ratio of confirmed AGNs to AGN candidates. Conversely, the contamination is defined as the number of confirmed non-AGNs divided by the number of AGN candidates. We classified as stars nine sources which, in both the discussed color vs. color diagrams (see Section 4.3), lie on the corresponding stellar locus, and which are classified as stars in the COSMOS Photometric Redshift Catalog (Ilbert et al., 2008); such subsample includes the source classified as a star on the basis of photometry. As a consequence, we can state that our secure sample includes 3% (9/265) contaminants. This leaves 36 sources in our secure sample with no classification and hence, in the

worst-case scenario where all of them are spurious candidates, contamination would rise to 17%. Nonetheless, we point out that 24 of the 36 non-classified sources lie on the edges of the field, very close to the edge areas that we masked or very close to other masked regions: it is thus very likely that their variability is just an effect of their position in bad regions that should have been masked, and as a consequence we should exclude them from our secure sample. We had already noticed these sources and their positions at the time we obtained our secure sample, but we did not exclude them *a priori* because we did not want to introduce a bias in the sample. Their exclusion would reduce the secure sample to 241 AGN candidates, of which we would hence confirm 91%, with 4% contaminants, which could rise to 9% if all of the remaining 12 non-classified sources were confirmed to be non-AGNs.

If we consider that the analyzed sky area, not including masked regions, corresponds to ≈ 0.80 square degrees, the number of confirmed AGNs returns a density of ≈ 275 AGNs per square degree.

From Table 6.1 we can see 32% of the 220 confirmed AGNs is validated by all the analyzed diagnostics, and the percentage is more than doubled –i.e., 74%– if we include the AGNs confirmed by 4/5 of the diagnostics. We note that there are 11 sources classified as AGNs and that do not have an X-ray counterpart; a star marks the corresponding lines in the table in order to ease their identification.

As for the VST-CDFS, our method combined to additional diagnostics allows to confirm 102 AGNs, 21 SNe, and 6 stars; if we exclude the SNe from the sample of 175 AGN candidates, we reduce such sample to 154 sources, and we can state that we confirm 66% (102/154) of them; this defines the purity of our sample, and means that, in the worst-case scenario, we would have 34% contaminants. Nevertheless, if we restrict our conclusions to the

sources for which information from at least one of the multi-wavelength CDFS catalogs is available, the sample is further reduced to 128 sources, and the purity rises to 80% (102/128). Such fraction is roughly consistent with our findings from the VST-COSMOS data analysis. A spectroscopic follow-up of the sources we did not confirm could help constraining the purity fraction.

Going back to the VST-COSMOS data analysis, in Section 4.1 we broached the issue of the increasing number of X-ray sources with an optically variable counterpart that we retrieve with respect to the total of X-ray emitters in the main sample. The fraction was 11% when *D0* observations over five months were used, while it rises to 32% when using the *D1* dataset, consisting of observations over a three year baseline. We can define the corresponding completeness as the number of X-ray sources that we confirmed as AGNs on the basis of their optical variability divided by the number of confirmed AGNs known *a priori* from the X-ray catalogs and having a VST counterpart: this was 15% when the analysis was limited to five month observations (De Cicco et al., 2015), and is 41% after the analysis of the *D1* dataset here presented. In De Cicco et al. (2015) we computed the completeness in three different magnitude bins of the same size from $r(\text{AB}) = 20$ mag to $r(\text{AB}) = 23$ mag, and obtained 26% for the 20 – 21 mag bin, 23% in the middle bin, and a drop to 5% in the bin corresponding to fainter magnitudes. Thus, not only the completeness was low, but it decreased when fainter objects were taken into account. The corresponding percentages obtained from *D1* data are 58%, 48%, and 30%, respectively, hence it is apparent that the completeness still drops at fainter magnitudes, but it is definitely higher in each bin due to the longer observing baseline.

The completeness for VST-CDFS data was estimated taking the sources falling in the AGN locus defined in Donley et al. (2012) as a reference, and

it is 22%. Such fraction is not far from the 15% completeness obtained from the five month analysis of COSMOS data, if we consider that the two were estimated on the basis of different reference samples of confirmed AGNs.

Figure 6.1 shows the location of all the confirmed AGNs in the VST-COSMOS field that are included in the X-ray catalogs and have a VST counterpart on a diagram like the one proposed in Figure 3.5, representing each source by means of its average magnitude and standard deviation; $D1$ and $D0$ data were used in the upper and lower panel, respectively. It is apparent that, in the case of $D0$ data, most of the X-ray sources are below the variability threshold; nevertheless, they have on average larger rms than the rest of the main sample population (the plots shown in the figure are limited to the two corresponding subsamples of non-variable objects). We verified that this still holds if we consider the three above mentioned magnitude bins. Such a result suggests that the optical variability of the majority of the X-ray sources is not detected mostly because of insufficient photometric accuracy.

We made use of a Kolmogorov-Smirnov (K-S) test to assess the extent of the average optical variability of the X-ray sources by comparing the rms of the subsample of non-variable objects in the VST sample to the rms of the non-variable X-ray sources having a VST counterpart and an X-ray luminosity $L_x > 10^{42} \text{ erg s}^{-1}$ (the luminosity threshold made us confident that the X-ray sample considered for the test consisted of active galaxies; see Section 4.1), and the test returned a probability $P \approx 10^{-7}$ that the two datasets were drawn from the same distribution, thus supporting the conclusions we had inferred from the plot.

All the performed tests show that large uncertainties can overshadow the optical variability of X-ray sources, especially the fainter ones, and this reflects into low values for the completeness with respect to the X-ray confirmed

AGNs; such an inconvenience could be overcome if the photometric accuracy were higher; lowering the variability threshold is also an option, although this would be at the expense of the purity of the sample of AGN candidates.

Trevese et al. (2008) perform an optical variability analysis of a set of eight observations of the CDFS over a 0.25 square degree area, imaged by the ESO/MPI 2.2 m telescope in La Silla; observations span about two years and data are from the Southern inTernediate Redshift ESO Supernova Search (STRESS) survey. The analysis is limited to the region with X-ray coverage, and the sample of AGN candidates consists of 104 sources. The obtained purity is $\approx 60\%$, although unconfirmed LLAGNs could lie among the remaining 40% of the candidates, and so this is to be intended as a lower limit; the completeness is 44% and is computed with respect to the X-ray sources in the field (with known spectra and X-ray luminosities L_x (2-8 keV) $> 10^{42}$ erg s $^{-1}$): this suggests the existence of a dependency of completeness on the baseline, which is something we also expect on the basis of the known trend of AGN variability, which is characterized by a red noise spectrum and an increasing structure function towards longer timescales (e.g., de Vries et al., 2005, and references therein). Thus, before *D1* data were available for our analysis, we tested such hypothesis by means of simulations involving the AGN structure function (e.g., Bauer et al. 2009 and references therein), and estimated that an extension of the baseline from five months to two years would result in an increase of the intrinsic variability by $\sim 50\%$. Assuming the X-ray detected population to be representative of the whole population, we calculated that increasing the intrinsic variability of the X-ray emitters by 50%, and improving our photometric accuracy (in Trevese et al. 2008 it is slightly higher than in our early VST data) would have brought $\sim 36\% \pm 3\%$ of the sources above our variability threshold; this is in agreement with the

completeness of $44\%^{+6\%}_{-9\%}$ (1σ binomial confidence limits) measured by Trevese et al. (2008).

The results from our simulations were confirmed by the use of *D1* data: in the upper panel of Figure 6.1, where data cover the already mentioned three year baseline, we can still see that the average rms of the X-ray sources is larger than the average rms of the whole VST main sample but now, while the variability threshold is roughly unchanged, there are many more X-ray emitters above the variability threshold (50% completeness). This confirms the dependence of detection efficiency on the sampled timescales; hence, a longer observing baseline leads to a higher number of detected variable sources and, correspondingly, to a higher completeness. This is also shown in Figure 6.2, where the fraction of X-ray confirmed AGNs retrieved for different baselines is represented.

On the basis of the illustrated results we can state that optical variability constitutes a robust criterion for AGN selection; this, especially when coupled with a higher photometric accuracy and a longer observing baseline, is a promising finding in the framework of current and future wide-field surveys (e.g., DES, LSST, see, e.g., Abbott & the DES collaboration 2016; Abell & the LSST collaboration 2009; Peters et al. 2015, and references therein), where variability studies will be crucial both for the discovery and the analysis of AGNs as well as other variable sources.

Table 6.1: Confirmed sources in our secure sample. We list the number of objects confirmed by each diagnostic, and also the number of AGNs (when it is not 0) confirmed by each combination of indicators.

confirmed sources (either AGNs or stars)	229 (86% of 265)
confirmed AGNs	220 (83% of 265)
confirmed stars	9 (3% of 265)
spectroscopic/SED validation (S)	202
X/O validation (X)	206
r - z vs. z - k color-color diagram validation (rzk)	173
MIR color-color diagram validation, Lacy region (L)	190
MIR color-color diagram validation, Donley region (D)	90
S+X+rzk+L+D validation	71
S+X+rzk+L validation	75
S+X+L+D validation	15
S+rzk+L+D validation	2
S+X+L validation	13
S+X+rzk validation	11
X+rzk+L validation	5
rzk+L+D validation*	1
S+X validation	12
rzk+L validation*	3
X+rzk validation	2
X+L validation	2
L+D validation*	1
only L validation*	3
only rzk validation*	3
only X validation	1
classified sources	
with no X-ray counterpart	11

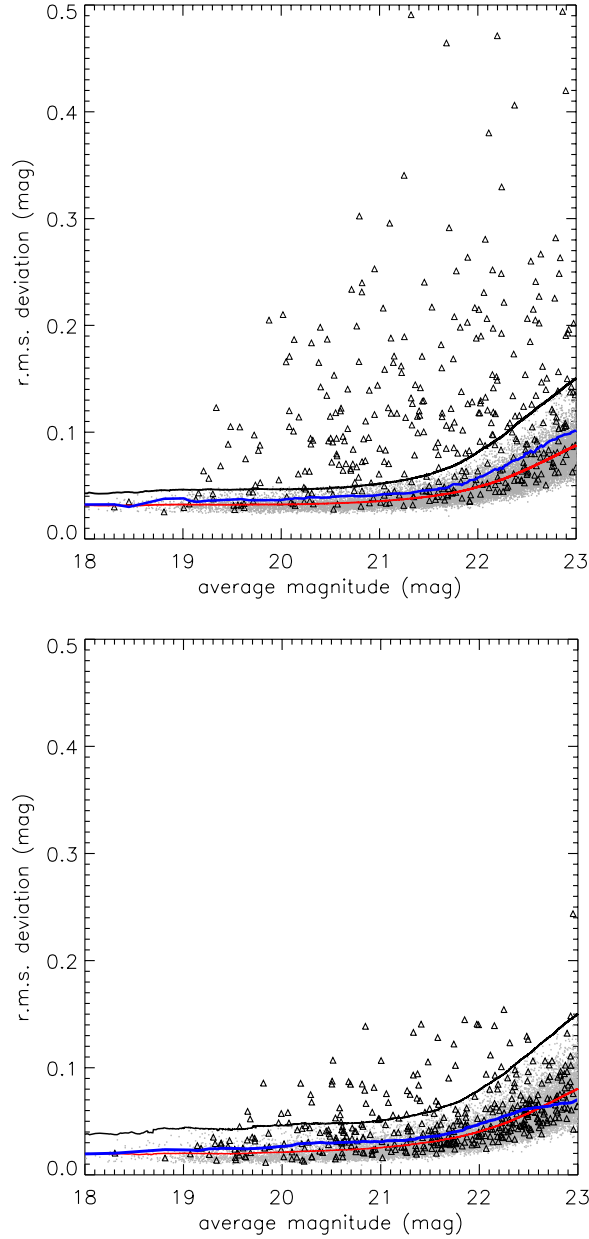


Figure 6.1: *Upper panel*: rms deviation from average magnitude as a function of average magnitude for the X-ray sources that are confirmed AGNs (triangles), from the *D1* dataset. The grey dots represent all the non-variable sources in the VST main sample. The running average of the rms deviation of the non-variable sources in the main sample (red line) and of the subsample of non-variable X-ray emitters (blue line), plus the variability threshold (black line) are also shown. 50% of the X-ray sources that are confirmed AGNs after X-ray catalogs are above the variability threshold. *Lower panel*: same plot as above, referring to the *D0* dataset. Here the fraction of X-ray confirmed AGNs above the variability threshold drops to 15%.

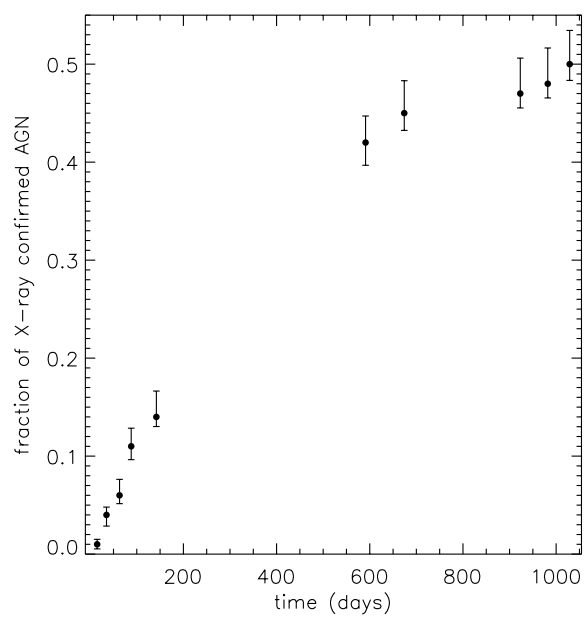


Figure 6.2: Fraction of AGNs confirmed after X-ray catalogs and retrieved through their optical variability as a function of the length of the baseline. Error bars are 1σ binomial confidence limits.

**Broad Absorption Line
Quasi-Stellar Objects:
Disappearance of the C IV Line
Trough in the SDSS-BOSS
Sample**

Chapter 7

Sample Selection and Data Reduction

7.1 The Baryon Oscillation Spectroscopic Survey

Our analysis of C IV BAL disappearance is based on data from BOSS, which is the largest of the four SDSS-III surveys (e.g., Eisenstein et al., 2011). BOSS was designed to map the baryon acoustic oscillation (BAO) signature imprinted in QSOs and luminous red galaxies (LRGs) by tracing their spatial distribution, aiming at measuring cosmic distances with the ultimate goal of defining improved constraints on the acceleration of the expansion rate of the Universe.

Observations were made with the 2.5 m SDSS telescope (Gunn et al., 2006) at Apache Point Observatory, New Mexico, USA. The telescope is equipped with two identical spectrographs, each with a camera for red and one for blue wavelengths, covering the wavelength ranges $5650 - 10000 \text{ \AA}$ and

3600 – 6350 Å, respectively. Spectra of numerous sources are acquired simultaneously by means of aluminium plates subtending a 3 degree-wide area on the sky and attached to the telescope; 1000 holes are drilled on each plate, in correspondence of the position on the sky of the sources to be observed, and a 2''-diameter fiber is plugged into each hole. Light from each observed source is directed to a beamsplitter covered with a dichroic coating that allows to split the red and blue components of the spectrum, so that each of them is registered on a different CCD. The spectral resolution varies within the range 1850 – 2650 in the red channel and 1560 – 2270 in the blue channel.

BOSS observations span four and a half years (fall 2009 – spring 2014) and survey more than 10000 square degrees previously monitored by SDSS-I/II surveys; approximately 1.5 million luminous galaxies with a redshift in the range 0.15 – 0.7, together with ≈ 150000 QSOs with $2.15 \leq z \leq 3.5$, were targeted.

7.2 Selection of QSOs and Spectra

SDSS-I/II allowed the identification of thousands of BAL QSOs (see Section 1.5). Gibson et al. (2009) presented a catalog of 5039 BAL QSOs, selected from the SDSS Data Release 5 (Schneider et al., 2007). The BOSS survey retargeted, among other sources, 2005 of such QSOs, aiming at the investigation of BAL variability on multi-year timescales, in order to gain knowledge of structure, dynamics and physical properties of QSO winds (Dawson et al., 2013). The selected QSOs fulfill the following requirements:

- are optically bright, i.e., have an i band magnitude < 19.3 mag;
- show strong BALs, i.e., have a balnicity index (following the definition by Gibson et al. 2009; see below) $BI_0 > 100 \text{ km s}^{-1}$ in at least one of

the observed BAL troughs;

- are characterized by a high signal to noise ratio (S/N; specifically, in Gibson et al. 2009 it is indicated as SN_{1700} , and it is required to be > 6 ; see below).

The balnicity index was proposed by Weymann et al. (1991) in order to quantify the “bal-ness” of a QSO by means of a continuous indicator characterizing the C IV absorption lines in a spectrum; it was defined as

$$- \int_{25000}^{3000} [1 - f(v)/0.9] C dv , \quad (7.1)$$

where $f(v)$ is the continuum-normalized flux as a function of the velocity displacement with respect to the center of the BAL line, while C is a constant that is set to 0 and switches to 1 only when the quantity in the brackets is continuously positive over a velocity range $\geq 2000 \text{ km s}^{-1}$. Such a definition allows to measure the EW^1 of an absorption line (in km s^{-1}) provided that the line is broader than 2000 km s^{-1} and is at least 10% below the continuum. The 25000 km s^{-1} blue limit and the 3000 km s^{-1} limit were set to avoid contamination from Si IV emission/absorption features on the blue end of the line and from the C IV emission line on the red end (see Figure 1.8, where the spectrum of one of our sources is shown and the wavelength range at issue can be seen). Gibson et al. (2009) defined a modified balnicity index BI_0 , where the red end limit is replaced by 0 km s^{-1} .

SN_{1700} was introduced in Gibson et al. (2009) to quantify the S/N corresponding to the measurement of C IV absorption; it is defined as the median

¹The EW is defined on the basis of the area of a spectral line below the continuum intensity level: the line is associated to a rectangle having the same area and, assuming that the intensity level I is the height of the rectangle, the division of the area by I returns the equivalent width of the line.

of the flux divided by the noise (provided by the SDSS pipeline), and is computed for the spectral bins in the wavelength range $1650 - 1750 \text{ \AA}$, which was chosen because it is close enough to the C IV BAL region and usually characterized by little absorption. Several factors (e.g., the integration time or the source redshift and luminosity) can affect the estimate of SN_{1700} ; nevertheless, it allows to select QSO samples on the basis of their S/N.

Our analysis focuses on C IV BALs since these are the most commonly observed ones, as $\approx 85\%$ of BAL QSOs are HiBALs (e.g., Wildy et al., 2014, and references therein; see also Section 1.5, where HiBALs were introduced), and hence exhibit C IV BALs; plus, contamination by adjacent absorption features is generally not relevant.

Basically, the disappearance of a BAL can be detected by comparing two or more spectra of the same QSO taken at different times; as a consequence, we needed sources with spectral coverage both from BOSS and previous SDSS-I/II surveys: this restricted our sample to 1606 out of 2005 QSOs, since there are no BOSS spectra for the rest of the sources. A further restriction was necessary in order to select the redshift window of interest: this was identified to be $1.68 < z < 4.93$ for C IV BALs to be fully visible in SDSS spectra, considering that their blueshifted velocities can range -30000 to 0 km s^{-1} (Gibson et al., 2009); as a result, the sample of QSOs was cut down to 1525 sources.

Following other works from the literature (e.g., Filiz Ak et al., 2012), we required our BALs to have a maximum velocity $-30000 \leq v_{\text{max}} \leq -3000 \text{ km s}^{-1}$ —and thus excluded all the BALs entirely confined in the range $-3000 - 0 \text{ km s}^{-1}$ —on the basis of what stated above about contamination by other features; in Section 8.2.1, when dealing with the v_{min} distribution of the BALs in our sample, we will show that the effect of our cut on the obtained

results is negligible.

Our final sample consists of all the QSOs whose SDSS-I/II spectra show at least one C IV BAL in the velocity range of interest, that is to say, 1319 sources (hereafter, main sample). For each QSO, at least a pair of spectra—one from SDSS-I/II and one from BOSS observations—is available; 343 sources have one or more additional spectra from SDSS-I/II and/or BOSS. Throughout the present work, unless otherwise stated, we will label as “pair of spectra” (or, equivalently, “pair of epochs”) a couple of spectra of the same source, where the less recent spectrum is from SDSS-I/II and the more recent one is from BOSS. The rest-frame timescales between observations in a pair range 0.28–4.9 yr. SDSS-I/II spectra cover the wavelength range 3800–9200 Å, while BOSS spectra cover bluer and redder wavelengths (3650–10400 Å).

7.3 Spectral Data Reduction

In the reduction process of BOSS spectra, we needed to take into account systematics mirroring spectrophotometric calibration errors (Margala et al., 2016). These mostly arise from the differential diffraction of light in the atmosphere and from a focal plane offset of the position of the fibers targeting QSOs with respect to the ones targeting standard stars used for calibration; as an effect, we get a higher throughput in the Lyman- α wavelength window when we observe high redshift QSOs, and this reflects into a higher S/N but, on the other hand, we get larger calibration errors with respect to those in SDSS-I/II spectra. We made use of the corrections presented and discussed in Margala et al. (2016) to correct such calibration errors.

In the header files of SDSS spectra there are bitmasks quantifying the “goodness” of every single pixel constituting the detector, based on a set

of observational or instrumental conditions. SDSS spectra are generally obtained by the combination of three or more exposures, and there may be pixels whose goodness changes from an exposure to another; pixels that are bad in all the exposures can be identified by means of the *and_mask* column in the Header Data Unit. Following Filiz Ak et al. (2012) and Grier et al. (2016), we masked all the pixels that in the header files are flagged as bad with respect to the “BRIGHTSKY” threshold, as this indicates that the flux contribution from the sky in such pixels is too high.

7.3.1 Extinction Correction

Spectra must be corrected for Galactic extinction before use. In Cardelli et al. (1989) an extinction law $A(\lambda)/A(V)$ is derived, where $A(\lambda)$ is the absolute extinction at the wavelength of interest and $A(V)$ is the absolute visual extinction; historically, the V -band is used as a reference. Such extinction law is to be used in the wavelength range $0.125 \leq \lambda \leq 3.5 \mu\text{m}$, and depends on the parameter $R_V = A_V/E(B - V)$, i.e., the ratio of visual extinction to reddening. We followed Cardelli et al. (1989) to correct for Galactic extinction, and adopted a Milky Way extinction model, where we assumed $R_V = 3.1$; visual extinction coefficients are from Schlegel et al. (1998). Once the correction was performed, we converted the observed wavelengths to rest frame wavelengths making use of redshifts from Hewett & Wild (2010).

7.4 Continuum Fit

We followed the steps to fit a continuum model to our spectra illustrated by Grier et al. (2016) and Gibson et al. (2009). We adopted a reddened power law model and made use of the Small Magellanic Cloud-like reddening curve

presented and discussed by Pei (1992).

In order to properly fit the continuum, it is necessary to identify a set of wavelength windows where typically there are no strong emission/absorption features: such windows are hence named relative line-free (RLF) regions. The same set of RLF regions was used for each spectrum, although in some cases one or more regions may be not available in a spectrum. The wavelength windows that we selected are $1280 - 1350 \text{ \AA}$, $1700 - 1800 \text{ \AA}$, $1950 - 2200 \text{ \AA}$, $2650 - 2710 \text{ \AA}$, and $3010 - 3700 \text{ \AA}$, and most of them were used in several works from the literature (e.g., Gibson et al., 2009; Filiz Ak et al., 2012); here, following Grier et al. (2016), we reduced the first region in the list on its blue side, in order to limit contamination from possible nearby emission features; furthermore, we introduced an additional RLF window corresponding to the wavelength range $1425 - 1450 \text{ \AA}$, to be used if a source has a redshift $z < 1.85$, in order to improve the blue end fitting of the corresponding spectrum.

We assigned a weight to each of the pixels in each RLF region on the basis of the total number of pixels constituting the region itself, in order to attribute the same weight to each region involved in the fitting process, regardless the number of pixels making up the region. In order to fit the continuum model to a spectrum, we performed a non-linear least square analysis iteratively, excluding outliers at each iteration on the basis of a 3σ threshold, thus minimizing the risk of contamination by prominent emission/absorption lines happening to fall in the selected RLF regions.

Uncertainties in the continuum fit were quantified by means of “flux randomization” Monte Carlo iterations (Peterson et al., 1998; Grier et al., 2016): essentially, we modified the flux in each spectral pixel by a random Gaussian deviate on the basis of the spectral uncertainty, then we fitted the continuum to the new spectrum we obtained, and iterated the procedure 100 times. We

assume the uncertainty of the continuum fit to be the standard deviation of the 100 iterations.

In Section 7.2 we mentioned that SDSS-I/II and BOSS spectra have different spectral coverage ($3800 - 9200 \text{ \AA}$ vs. $3650 - 10400 \text{ \AA}$, respectively); as a consequence, we cropped BOSS spectra adjusting them based on the SDSS-I/II wavelength window, thus ensuring that the same RLF regions were used in the continuum fit process when dealing with multiple spectra of a same source.

In Figure 7.1 we report an example of a source where disappearing C IV BAL troughs are observed; RLF regions and best-fit continuum model are also shown, together with the main emission features that are typical of BAL QSOs in the observed wavelength range.

SDSS spectra are identified by three integer numbers:

- plate, which indicates the aluminium plate used to obtain the spectrum;
- MJD, which stands for Modified Julian Date, and thus refers to the night of the observation;
- fiber, which identifies the fiber used to observe that specific object; this is a number in the range $1 - 1000$ for BOSS observations, and in the range $1 - 640$ for SDSS-I/II spectra, because a reduced number of fibers was used at the time.

Plate, MJD, and fiber numbers are reported on top of each panel in Figure 7.1 in order to allow the identification of the corresponding sources.

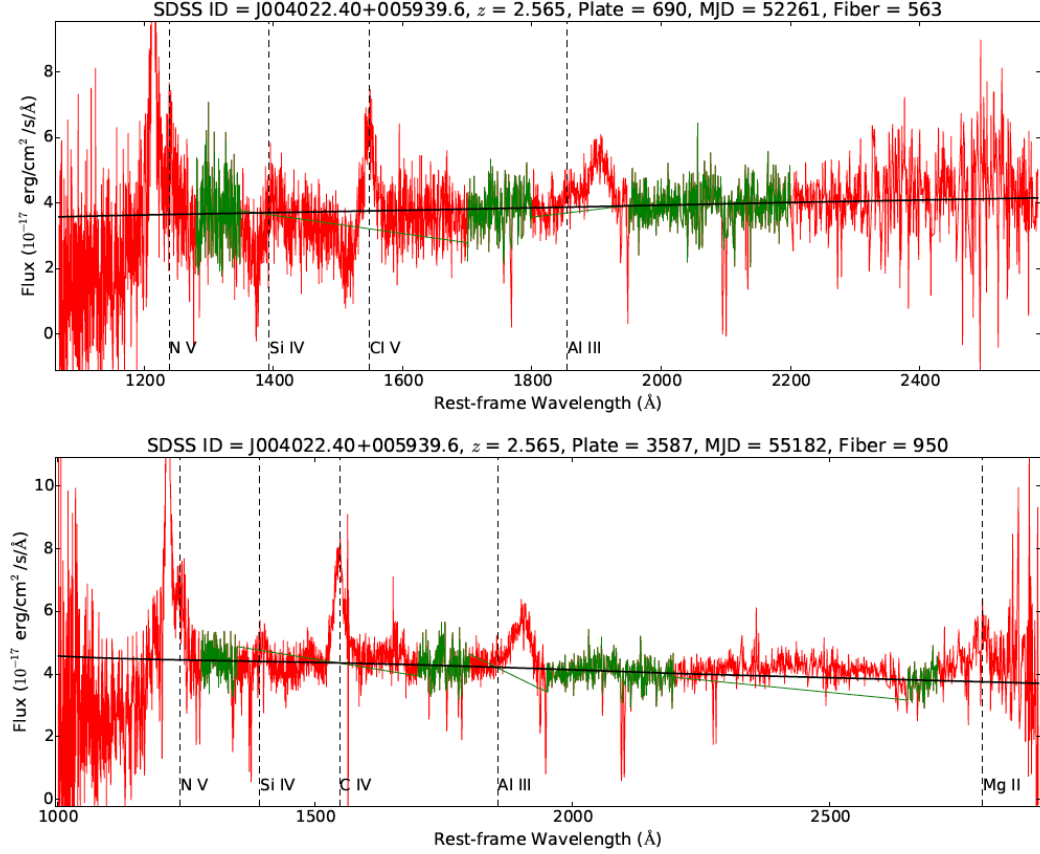


Figure 7.1: Examples of continuum model fits for a pair of spectra (SDSS-I/II in the upper panel, BOSS in the lower panel) where we observe C IV BAL disappearance. The red line represents the flux as a function of the rest frame wavelength, while the green regions stand for the RLF windows used to fit the continuum model, shown by the solid black line. Vertical dashed lines indicate rest frame wavelengths corresponding to features that are typically seen in QSOs. The source ID in the SDSS archive, together with its redshift and information about the plate, MJD, and fiber corresponding to the observation, are also reported on top of each panel.

Chapter 8

Statistical Analysis of Disappearing BAL Troughs

8.1 Identification of Disappearing BAL Troughs

The analysis here illustrated aims at investigating the disappearance of C IV BAL troughs in the largest sample of BAL QSOs ever available so far; the existence of a coordination in the variability of multiple troughs corresponding to the same transition is also investigated. The size of the sample allows to perform a reliable statistical analysis, with the ultimate goal of shedding light onto the physical processes originating BAL variability and onto the structure and geometry of the region where winds form and propagate.

In order to facilitate the identification of BAL troughs, we smoothed each spectrum making use of a three pixel-wide boxcar algorithm which, essentially, consists in performing a running average over three pixel-wide bins. We converted wavelengths into velocities by means of redshifts and identified all the C IV BAL troughs in each SDSS-I/II spectrum: for a trough to be included in our sample we required its width in terms of velocity to

be $\Delta v \geq 2000 \text{ km s}^{-1}$ and its flux to extend below 90% of the normalized continuum level. Such selection criteria returned a sample of 1874 BAL troughs, detected in the spectra of the 1319 sources constituting our main sample; this means that, in some cases, in a spectrum we detected more than one C IV BAL trough, corresponding to different velocities.

In the BOSS spectrum/spectra associated to each of the sources in the main sample, we examined the same wavelength window where the BAL in the SDSS-I/II spectrum was detected. We assumed that a BAL detected in a spectrum from SDSS-I/II disappeared if in the corresponding BOSS spectrum we do not see any absorption line extending below 90% of the normalized continuum level at all, or if the BAL turns into a NAL, i.e., it has a width $\Delta v < 500 \text{ km s}^{-1}$.

Associating a BAL to another when comparing two different spectra is not always trivial, since it can happen that one of the troughs is shifted with respect to the other; we assumed there is mutual correspondence between two troughs if they cover wavelength ranges that overlap at least partially.

Our inspection returned a sample of 105 disappearing C IV BAL troughs in the spectra of 94 sources. For some of these 94 sources, more than a spectrum pair was available; we therefore decided, when dealing with such cases, to consider the most recent among the SDSS-I/II spectra where a BAL trough is seen and the least recent BOSS spectrum where the BAL disappears, in order to probe the shortest accessible timescales and the fastest variability.

A criterion assessing the significance of the observed BAL disappearances is necessary in order to minimize contamination from spurious disappearances. Following Filiz Ak et al. (2012), we made use of a two-sample χ^2 test performed on the two sets of data points corresponding to the flux in each

pair of wavelength windows where we observe a disappearance, and we required the probability P_{χ^2} associated to the test to be $\leq 10^{-8}$ for the change in a trough to be considered as random; hence, if $P_{\chi^2} \leq 10^{-8}$, we can discard the null hypothesis and be confident that the observed disappearance is real.

Based on the defined threshold, we obtained a sample of 56 disappearing BAL troughs detected in 52 sources (hereafter, P_8 sample). Nevertheless, a visual inspection of each of our disappearing BAL candidates revealed that a number of excluded disappearances may in fact be real, suggesting that, while returning a highly reliable sample, our threshold might be over-conservative. As a consequence, we selected a second sample of disappearances that looked reliable on the basis of visual inspection, and such a sample turned out to correspond to troughs with a probability $P_{\chi^2} \leq 10^{-4}$ for disappearances to be accidental. The new selection returned 17 additional disappearing BALs observed in 16 sources; we shall refer to the full sample of sources for which $P_{\chi^2} \leq 10^{-4}$ as P_4 sample. In total, this consists of 73 (56+17) disappearing BAL troughs detected in the spectra of 67 sources. There is one source belonging to both subsamples, as it shows two disappearing BALs with $P_{\chi^2} \leq 10^{-8}$ and one disappearing BAL with $10^{-8} \leq P_{\chi^2} \leq 10^{-4}$; this explains why the sum of the sources in the P_4 sample is 67 instead of 68. In what follows we will generally report the results of our analysis for the P_4 sample, but we will also discuss some relevant results concerning the P_8 sample, which is to be considered more reliable; this will also allow us a proper comparison between our findings and those from Filiz Ak et al. (2012). In Table 8.1 numerical details about the main sample as well as the P_4 and P_8 samples are reported.

In Figure 8.1 we report some examples of spectrum pairs where C IV BAL disappearance is detected. It is apparent that, in some cases, the disappear-

Table 8.1: Detailed information about our main sample of sources, the P_4 and P_8 samples, and the C IV BAL troughs in their spectra. We remind the reader that the main sample consists of all the sources for which at least a pair of spectra (one from SDSS-I/II and one from BOSS) is available and where at least one C IV BAL trough is detected in the SDSS-I/II spectrum/spectra.

<i>MAIN SAMPLE</i>	
SDSS-I/II spectra	1543
BOSS spectra	1654
SDSS-I/II spectra exhibiting C IV BAL troughs	1319
C IV BAL troughs detected in SDSS-I/II spectra	1874
<i>P₄ SAMPLE</i>	
Sources with a disappearing BAL trough	67
Disappearing BAL troughs in BOSS spectra	73
<i>P₈ SAMPLE</i>	
Sources with a disappearing BAL trough	52
Disappearing BAL troughs in BOSS spectra	56

ing BALs are more than one, and there may also be additional BAL troughs that do not disappear; this is something we will present in Section 8.2.2.

The redshift distributions for all the sources in the main sample and in the P_4 sample are presented in Figure 8.2, while the rest-frame time difference

Δt between the two spectra in a pair where disappearance is observed is shown in Figure 8.3 for all the sources in the P_4 sample (this may sound redundant, but there are some of the sources in the P_4 sample for which more than two spectra are available, and disappearance is not necessarily observed in each pair). We already mentioned that the rest-frame timescale between observation pairs is in the range $0.28 - 4.9$ yr; the average Δt is ≈ 1123 days, while the median Δt is ≈ 1146 days, and both indicate that, on average, we observe disappearances on a timescale of ≈ 3.1 yr.

8.2 Statistical Properties of Disappearing BALs

So far we defined a main sample of 1319 sources where 1874 C IV BAL troughs have been detected, and we introduced the P_4 sample, consisting of 73 disappearing BAL troughs, detected in the spectra of 67 sources; we also identified a more reliable sample of 56 disappearing BALs (P_8 sample), observed in the spectra of 52 sources.

On the basis of our findings, we can estimate the average lifetime of a BAL trough and of the BAL phase along our line of sight, in order to gain global insight into the BAL phenomenon over large timescales. We computed the fraction of disappearing BAL troughs $f_{\text{disapp}_{P_4}} = 73/1874 = 3.9^{+0.5}_{-0.5}\%$ and the fraction of QSOs exhibiting at least one disappearing BAL trough in their spectra $f_{\text{QSO}_{P_4}} = 67/1319 = 5.1^{+0.7}_{-0.6}\%$ (error bars on percentages are computed following Gehrels 1986, where approximated formulae for confidence limits are derived assuming Poisson and binomial statistics). The two fractions become $f_{\text{disapp}_{P_8}} = 56/1874 = 3.0^{+0.5}_{-0.4}\%$ and $f_{\text{QSO}_{P_8}} = 52/1319 = 3.9^{+0.6}_{-0.5}\%$, respectively, if we restrict our analysis to the P_8 sample. Despite the errors on fluxes, which are on average on the order of 10^{-2} , the high reliability of

the disappearing BAL selection is a consequence of the probability threshold we chose for the χ^2 test.

The estimated disappearance frequency allows us to estimate the average rest-frame lifetime $\bar{t}_{\text{trough}_{P_4}}$ of a BAL trough along our line of sight: we can roughly define it as the average value of the maximum time difference $\langle \Delta t_{\text{max}_{P_4}} \rangle$ between two epochs in our main sample divided by the fraction of BAL troughs $f_{\text{disapp}_{P_4}}$ that disappear over such time; since $\langle \Delta t_{\text{max}_{P_4}} \rangle \approx 1144$ days, corresponding to ≈ 3.1 yr, we obtain $\bar{t}_{\text{trough}_{P_4}} \approx \langle \Delta t_{\text{max}_{P_4}} \rangle / f_{\text{disapp}_{P_4}} = 80^{+10}_{-10}$ yr¹. Limited to the BAL troughs in the P_8 sample, we obtain an average rest-frame lifetime $\bar{t}_{\text{trough}_{P_8}} \approx \langle \Delta t_{\text{max}_{P_8}} \rangle / f_{\text{disapp}_{P_8}} = 104^{+17}_{-14}$ yr, the average value of the maximum time difference being $\langle \Delta t_{\text{max}_{P_8}} \rangle \approx 1142$ days.

On the basis of several works from the literature (e.g., Hall et al., 2002; Filiz Ak et al., 2012) we can state that, if a source is a BAL QSO, BALs originating from C IV transitions are generally present in its spectrum, and they are typically the strongest that we observe. This means that, when all the C IV BAL troughs disappear from a spectrum, generally there are no other BALs left, meaning that we do not find any N V, nor Si IV, nor Ly α BALs², hence the source turns into a non-BAL QSO. Our P_4 sample includes 30 sources that change into non-BAL QSOs; we can therefore derive the fraction of QSOs turning into non-BAL QSOs as the ratio of such number to the total number of objects in our secure sample, that is, $f_{\text{transform}_{P_4}} = 30/1319 = 2.3^{+0.5}_{-0.4}\%$.

¹Here and in what follows we adopted different notations to make a distinction between average quantities that we directly measure, e.g., $\langle \Delta t \rangle$, and average quantities which we derive, e.g., \bar{t} .

²We note that, if no BALs corresponding to high ionization transitions are observed, lower ionization BALs will not be observed as well (see Section 1.5, where BALs are classified depending on the observed transitions).

Once we know such fraction, we can estimate the lifetime of the BAL phase in a QSO, which we can roughly define as the average of the maximum time difference (already used above) divided by the fraction of BALs that turn into non-BALs over that time range, i.e., $\bar{t}_{\text{BAL}} \approx \langle \Delta t_{\text{max}} \rangle / f_{\text{transform}}$; again, the estimate we obtain is limited to what we can see along our line of sight, as the only information we have concerns what happens in this direction, and the disappearance of a BAL does not necessarily mean that it does not exist anymore, as it could have simply moved out of our line of sight. For the P_4 sample we get $\bar{t}_{\text{BAL}_{P_4}} \approx 136^{+30}_{-24}$ yr; if we focus on the sources in the P_8 sample, the number of BAL QSOs turning into non-BAL QSOs reduces to 24, the corresponding fraction becomes $f_{\text{transform}_{P_8}} = 24/1319 = 1.8^{+0.5\%}_{-0.4\%}$, and the lifetime of the BAL phase along the line of sight is $\bar{t}_{\text{BAL}_{P_8}} \approx 174^{+48}_{-39}$ yr.

When dealing with such estimates, one should keep in mind that, even though all the BALs disappear from the spectrum of a source, other BALs can emerge at a later time, either in that same region or in a different one; as a consequence, the definition of “BAL phase”, as well as the resulting \bar{t}_{BAL} , should be handled with caution.

8.2.1 Velocity Distributions

Relevant information about the BAL trough population can be inferred from the analysis of the BAL properties in terms of velocity. In Section 1.5 we introduced the maximum and minimum velocity of a BAL trough, v_{max} and v_{min} , defined by the two ends of the absorption feature, the velocity difference $\Delta v = |v_{\text{max}} - v_{\text{min}}|$, and the central velocity v_c , which is the average of v_{max} and v_{min} .

In Figure 8.4 we compare the population of sources in the P_4 sample

to the sources in the main sample. Specifically, in panels (a) and (b) the v_{\max} and v_{\min} distributions for both populations are shown, respectively; we performed a K-S test on each pair of distributions (full results are reported in the various panels) in order to assess the probability that the two datasets are consistent. As for v_{\max} , we obtain that the maximum distance between the two cumulative distributions is $D \approx 0.18$, and the probability to get a higher value for D assuming that the two datasets are drawn from the same distribution function is $P \sim 10^{-2}$; as a consequence, we cannot state that the two distributions are inconsistent.

Different results are obtained with the v_{\min} distributions, as it is apparent that the bulk of the minimum velocities for the P_4 sample is clustered around higher values if compared to the corresponding distribution derived for the main sample. In this case we get a maximum distance $D \approx 0.34$ and a probability $P \sim 10^{-7}$, suggesting that inconsistency is plausible. This makes us confident that our choice to exclude from our analysis the BALs entirely confined in the velocity range $-3000 - 0 \text{ km s}^{-1}$ (see Section 7.2) does not affect significantly our results, since the v_{\min} distributions suggest that disappearing BALs are generally characterized by high values of their v_{\min} .

In panel (c) the central velocity distributions for the two samples are displayed: they resemble the distributions from panel (b), and we can see that the disappearing BAL troughs have a higher central velocity than the BALs in the main sample. We notice that the probability in this case is higher than in panel (b): this happens because the central velocity is affected not only by v_{\min} , but also by v_{\max} values. In panel (d) the velocity difference distributions are compared, and we deduce that disappearing BAL troughs are generally narrower than the ones in the main sample.

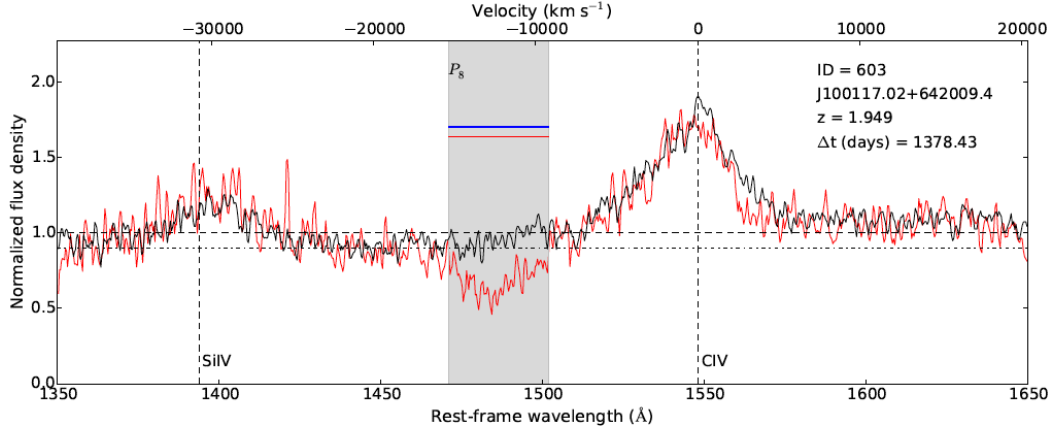


Figure 8.1: Overplot of SDSS-I/II (red) and BOSS (black) spectra showing disappearing BAL troughs. On top right of each panel we report the ID of the corresponding source from our catalog, the SDSS ID, the redshift, and the rest-frame time difference between the two represented spectra. Plots are restricted to the wavelength range $1350 - 1650 \text{ \AA}$ for a better visualization of the window where C IV BAL disappearance can be observed. Rest-frame wavelength (bottom) and velocity (top) are reported on the horizontal axes, while the normalized flux density is shown on the vertical axis. The horizontal dashed line represents the level where normalized flux density equals unity, while the dash-dot line indicates the 0.9 level for the normalized flux density: our analysis concerns BAL troughs extending below this threshold. The two vertical dashed lines represent the rest-frame wavelengths corresponding to the Si IV and C IV emission lines (1394 \AA and 1549 \AA , respectively). Red and black horizontal lines identify BAL troughs in the SDSS-I/II spectra and BOSS spectra, respectively, while blue bars mark SDSS-I/II BAL troughs that disappear in BOSS spectra. The lack of either black lines or blue bars in a BOSS spectrum in correspondence of a BAL trough in the SDSS-I/II spectrum means that the BAL trough has turned into one or more mini-BALs, or that the disappearance is not considered reliable (that is, $P_{\chi^2} > 10^{-4}$). The regions corresponding to SDSS-I/II BAL troughs are shaded for a better visualization. The writing “ P_8 ” on top of a shaded area marks disappearing BAL troughs belonging to the P_8 sample. In some cases, BAL troughs outside the velocity range $-30000 - 0 \text{ km s}^{-1}$ are apparent, but we remind that they were not taken into account in the present analysis (see Section 7.2).

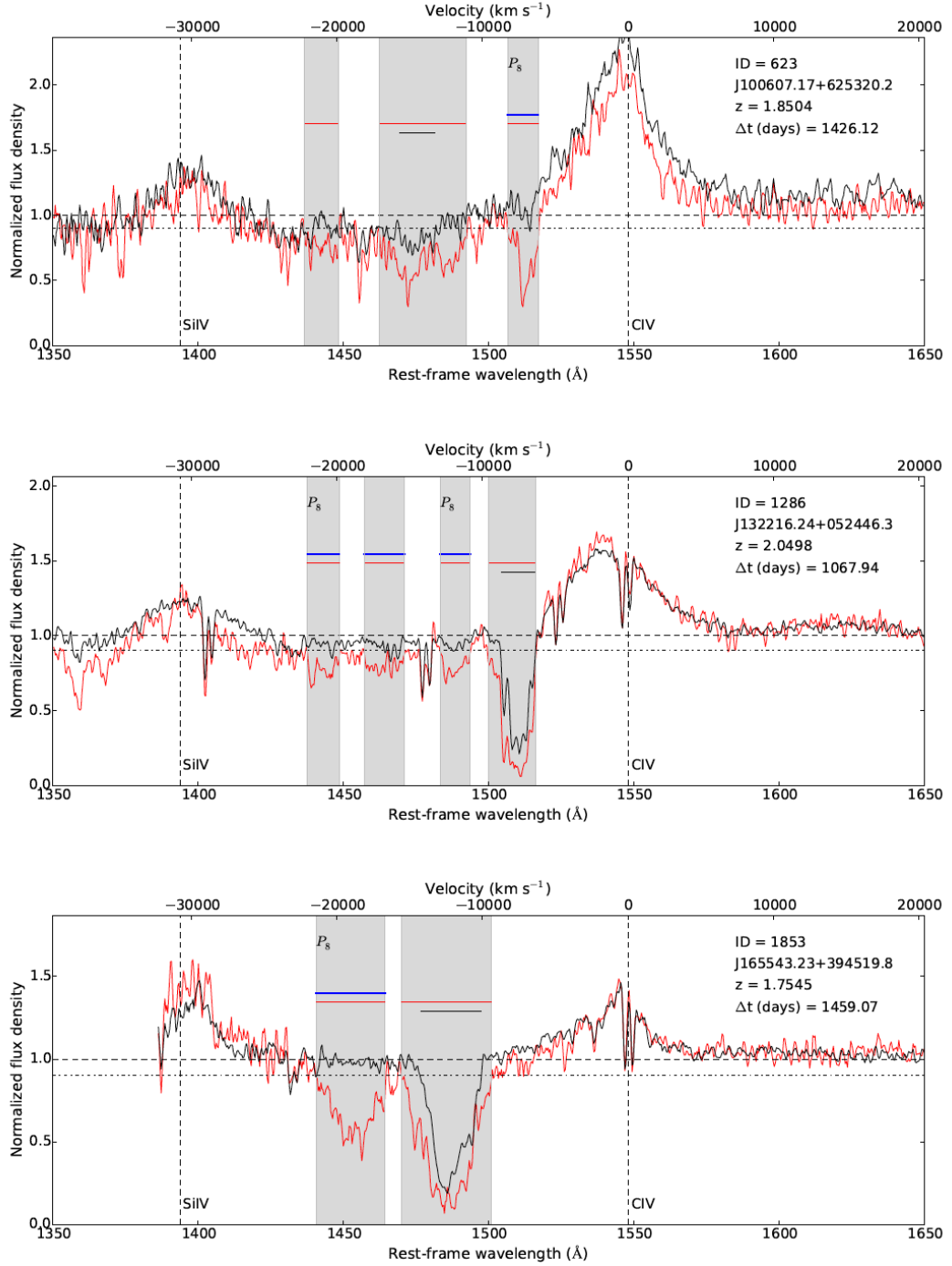


Figure 8.1: Overplot of SDSS-I/II (red) and BOSS (black) spectra showing disappearing BAL troughs (cont.).

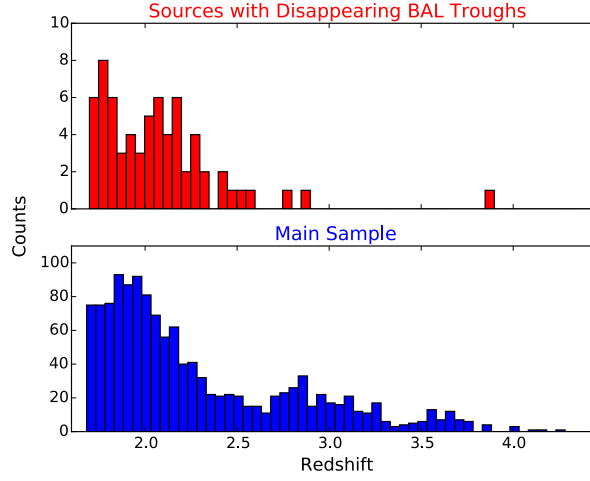


Figure 8.2: Redshift distribution for the sources in the P_4 sample (*top*) and in the main sample (*bottom*). Redshifts are from Hewett & Wild (2010).

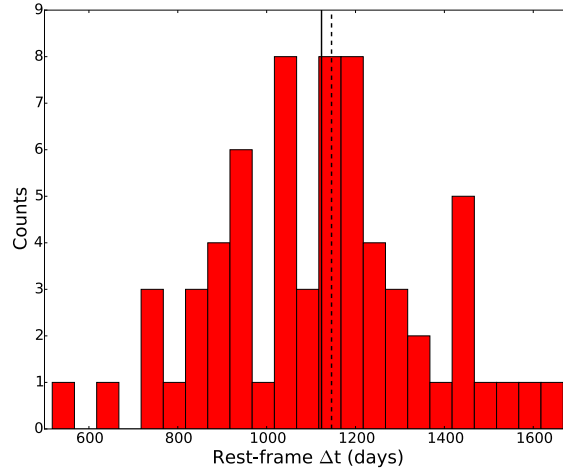


Figure 8.3: Distribution of the rest-frame time difference Δt between epoch pairs where disappearance is observed, for each source in the P_4 sample. The average (solid line) and median (dashed line) values for the time difference are shown.

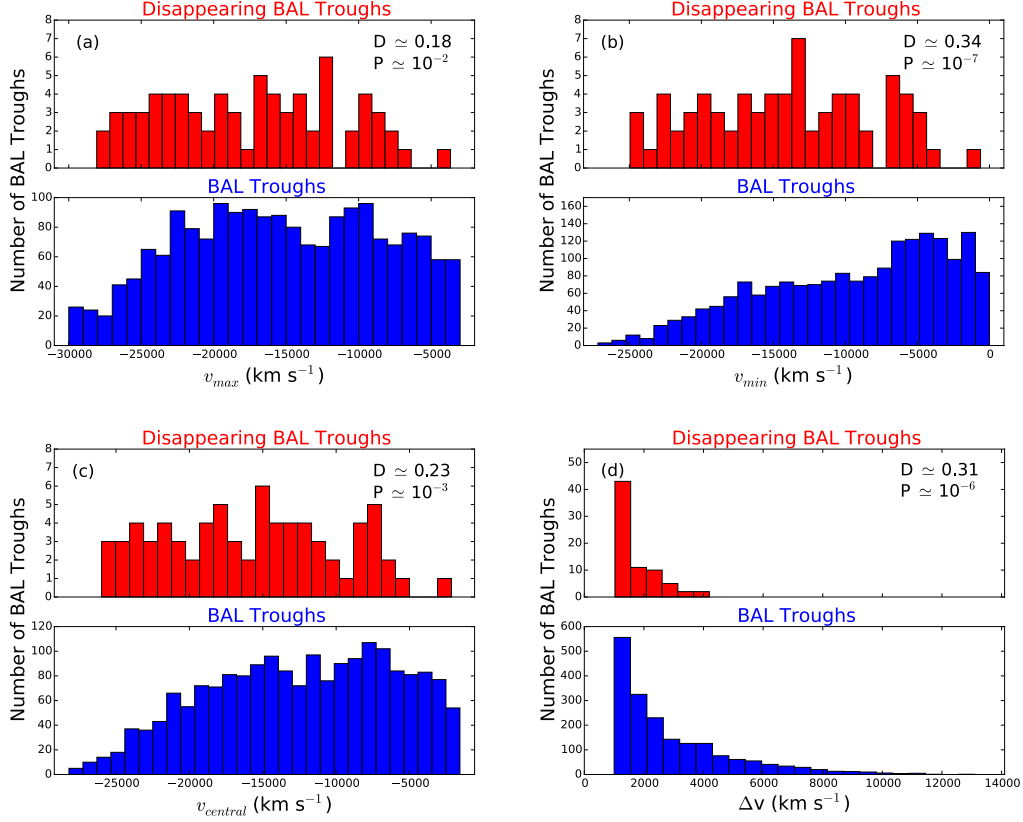


Figure 8.4: Maximum observed velocity v_{\max} (a), minimum observed velocity v_{\min} (b), central velocity v_c (c), and BAL width Δv (d) distributions for the P_4 sample (upper histogram in each panel) and for the main sample of sources with BAL troughs (lower histogram in each panel). Results of the K-S test performed on each pair of cumulative distributions are reported in each panel: D is the maximum distance between the two cumulative distributions, and P is the probability to get a higher D value assuming that the two datasets are drawn from the same distribution function. The corresponding distributions for the P_8 sample are shown in Figure 8.9, while the distributions obtained in the work by Filiz Ak et al. (2012) are presented in Figure 8.10, to allow the comparison between the two samples.

8.2.2 Equivalent Widths and Coordination in BAL Variability

In Section 8.1 we mentioned that, in some cases, a spectrum exhibits more than a C IV BAL trough, and not all of them disappear. When comparing two epochs in a pair, we can find BAL troughs in the less recent spectrum that are still BALs in the more recent spectrum, or BALs that turn into one or more mini-BALs; in this last case, on the basis of the definition of disappearance we stated in Section 1.5, we cannot state that a BAL disappears. It can also happen that, regardless the number of disappearing/non disappearing BALs, other BALs emerge in the more recent spectrum. Some of the listed instances were shown in Figure 8.1. The presence of additional non-disappearing BAL troughs in spectra where disappearances are detected is a chance for us to investigate the existence of a correlation in the variability of different BAL troughs when we compare two epochs of the same source.

The inspection of the 67 pairs of spectra corresponding to the sources in our P_4 sample revealed that there are 28 additional non-disappearing BALs in the spectra of 27 out of 67 sources. We chose not to take into account BALs turning into mini-BALs, and focused on BALs in the SDSS-I/II spectra that are still BALs in the corresponding BOSS spectra. Hereafter we shall refer to the subsample of the 28 additional non-disappearing BAL troughs (or, equivalently, to the subsample of the corresponding 27 QSOs) as *ND* sample.

Aiming at a further characterization of our P_4 and *ND* samples, we computed the EW (see Section 7.2) of each BAL trough: the calculation consists in integrating the quantity $1 - F_{\text{norm}}$ (where F_{norm} is the normalized flux) over wavelengths (or velocities); specifically, we summed all the areas $(1 - F_{\text{norm}}) \times d\lambda$ (or $\times dv$), where $d\lambda$ and dv represent the wavelength or veloc-

ity bins defined by the distance between two adjacent points in a spectrum.

In Figure 8.5 we report the EW distributions for all the BAL troughs in the main sample and in the ND and P_4 samples, in order to compare them. EW measures were always performed in the latest SDSS-I/II spectrum in the case of QSOs where we observe a disappearance and, in order to be consistent, they were performed in the most recent SDSS-I/II epoch as well for the rest of the QSOs belonging to the main sample, which were used as a reference (see further).

Once more, we made use of a K-S test to compare EW cumulative distributions. The probability of consistency for the main sample– P_4 sample pair is $P \sim 10^{-15}$ and the maximum distance is $D \approx 0.50$, while we get $P \approx 0.006$ when comparing the main sample to the ND sample, for which the maximum distance is $D \approx 0.32$ and so, in this last case, we cannot make any assumptions about inconsistency. We also notice that the disappearing BAL troughs are generally characterized by low EW values, the highest one being $< 17 \text{ \AA}$, while non-disappearing BAL troughs in the main sample typically reach much higher values ($\approx 80 \text{ \AA}$) for their EWs.

The comparison of the EWs of different BAL troughs in a pair of spectra allows to investigate the possible existence of a coordination in the variability of such BALs. The EWs of the non-disappearing BAL troughs in our ND sample are compared in Figure 8.6; the two epochs are always the ones where we observe disappearing BALs (i.e., the most recent SDSS-I/II epoch where we detect a BAL trough and the least recent BOSS epoch where that same BAL is no longer detectable). All the non-disappearing BAL troughs detected in the spectra of QSOs belonging to the main sample are also represented as a reference; in this case, the two epochs chosen for the comparison are the most recent among SDSS-I/II spectra and the least recent among

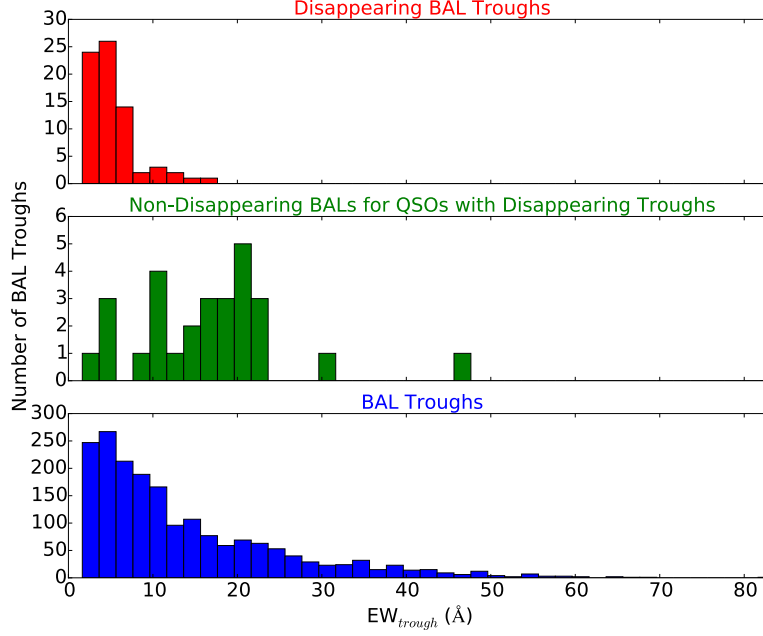


Figure 8.5: EW distributions for C IV BAL troughs in the P_4 sample (*upper panel*), ND sample (*middle panel*), and main sample (*lower panel*). EW measures for each BAL are from the latest SDSS-I/II epoch.

BOSS spectra.

The plot clearly shows that the distribution of the main sample BALs is roughly symmetrical on the two sides of the bisector (i.e., the line where the two EWs for a trough are equal); this indicates the absence of a dominant trend: BALs can get stronger as well as weaker. On the other hand, there are 22 out of 28 (79%) BAL troughs from the ND sample that weaken from the older epoch to the newer one: this means that, when in a spectrum there is more than one BAL trough and one of them disappears, in 79% of the instances the EW of the remaining BALs decreases when comparing the less recent to the more recent spectrum.

As a further test, we analyzed the trend of the fractional EW variation

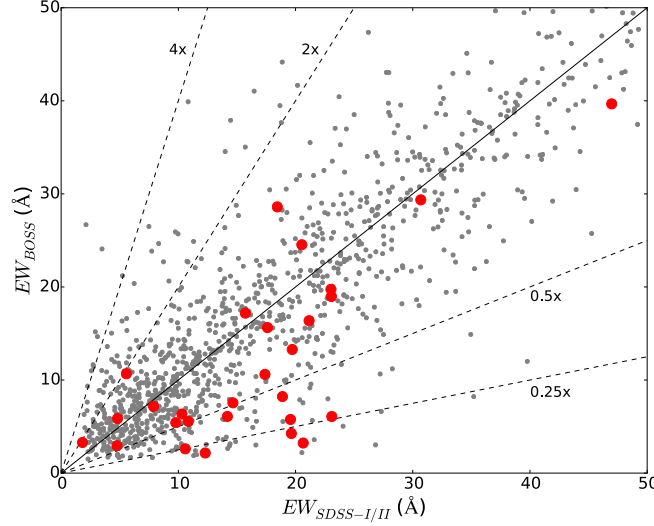


Figure 8.6: EWs at two different epochs for the QSOs belonging to the *ND* sample (i.e., showing both disappearing and non-disappearing BAL troughs in their spectra; large red dots), consisting of 27 sources where we detect 28 non-disappearing BAL troughs. For each BAL trough, the two EW measurements are obtained from the same epoch pair where disappearance is detected. Small grey dots in the background show the EWs at two epochs (one from SDSS-I/II and one from BOSS) for all the sources in the main sample having non-disappearing BAL troughs. In this case we chose the most recent SDSS-I/II epoch and the least recent BOSS epoch for each source. We limited our investigation to BAL troughs in SDSS-I/II spectra that correspond to BAL troughs in BOSS spectra, hence BALs turning into mini-BALs, or *vice versa*, are not taken into account. The solid line indicates where the EWs of the two compared epochs are equal, while the dashed lines indicate where the EW of the BOSS epoch is four times, two times, half of, and a quarter of the EW in the SDSS-I/II epoch.

for the non-disappearing BAL troughs in the *ND* sample: we define it as $\Delta EW / \langle EW \rangle$, where $\Delta EW = EW_{\text{BOSS}} - EW_{\text{SDSS-I/II}}$ and the average $\langle EW \rangle$ is computed over the two epochs. In Figure 8.7 the fractional EW variation is represented as a function of the offset $v_c - v_{c_{ND}}$ between the central velocity $v_{c_{ND}}$ of the non-disappearing BAL and the central velocity v_c of the disap-

pearing BAL detected in the same pair of epochs; where there is more than one disappearing BAL, the offset is computed with respect to the one having the highest central velocity. We found that for 27 out of 28 (96%) non-disappearing BAL troughs the velocity offset is negative: because of the way it is defined, and since we are dealing with blueshifted velocities, this means that in 96% of the instances in our *ND* sample the BAL trough that disappears is the one with the highest central velocity. Moreover, the already mentioned weakening trend is apparent: only six of the non-disappearing BALs (21% of 28) are characterized by positive values of the fractional EW variation, corresponding to EWs that are stronger in the BOSS epoch than in the SDSS-I/II epoch. We notice that the BAL troughs with a positive velocity offset (i.e., those with $v_{cND} < v_c$) generally weaken. The weakening trend is also observed in the BALs with the largest velocity offsets; all this probes the existence of a coordination in BAL trough variability and also suggests this is a persistent phenomenon.

8.3 Comparison to Results from Filiz Ak et al. (2012)

In Section 1.5 we mentioned that part of the spectra that we analyzed in our work was also inspected by Filiz Ak et al. (2012); a comparison of the findings is therefore presented here.

First, we cross-matched our main sample to the sample of QSOs examined in Filiz Ak et al. (2012): observations in the latter case have $\text{MJD} \leq 55811$, and the sample consists of 582 sources where 925 C IV BAL troughs are identified; the corresponding sample of disappearing BALs consists of 21 troughs detected in the spectra of 19 QSOs (hereafter, *FA* sample). The

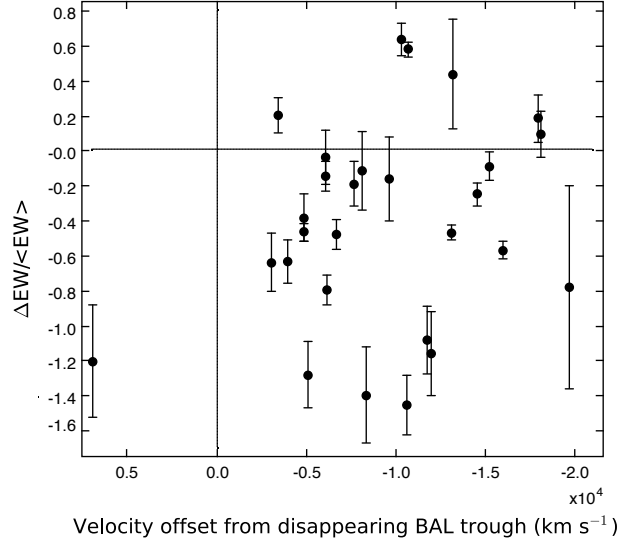


Figure 8.7: Fractional EW variations for the 28 non-disappearing BAL troughs in the *ND* sample. The two epochs used for each trough are the same as in Figure 8.6. The horizontal axis shows the offset of the central velocity $v_{c_{ND}}$ of each non-disappearing trough with respect to the v_c of the corresponding disappearing trough. The central velocity for a trough is defined as $v_c = (v_{max} + v_{min})/2$; error bars are computed propagating the errors on the EWs in the two epochs. Uncertainties on the EWs, in turn, are computed in an analogous way as errors on continuum fits (see Section 7.4), iterating the procedure 1000 times. The horizontal and the vertical line represent the zero level for each axis, and are plotted for a better identification of the various regions in the diagram. The estimate of error bars on this plot is preliminary. We invite the reader to refer to De Cicco et al. (in prep.) for the final version of the plot.

cross-match of our main sample with the *FA* sample returned 558 out of 582 sources; here we list the reasons why we did not include the remaining 24 objects in our main sample:

- six sources were excluded as no BOSS spectra were available for them;
- five sources were excluded because of problems in the reduction process;
- five sources were excluded since their SDSS-I/II spectra exhibit C IV

BAL troughs outside the velocity range of interest (all of them have $v_{\max} > -3000 \text{ km s}^{-1}$);

- eight sources were excluded since their SDSS-I/II spectra do not show C IV BAL troughs in the velocity range of interest, nor in the proximity of it. Nonetheless, an accurate inspection of the spectra at issue revealed that all of them show at least one mini-BAL trough in the velocity range of interest; each mini-BAL has a width $\Delta v > 1500 \text{ km s}^{-1}$ and, in particular, half of them have $\Delta v > 1920 \text{ km s}^{-1}$. This suggests that we do not detect the expected BALs due to slight differences in the spectrum fitting/normalization in the two works.

In our P_4 sample we retrieved 16 of the 21 disappearing BAL troughs constituting the FA sample, hence there should be five disappearances that we do not detect. A detailed analysis of the spectra of the corresponding sources showed that:

- the BOSS spectrum of J074650.59+182028.7 exhibits two non-deblended NAL doublets in the wavelength region where the BAL from the SDSS-I/II spectrum is supposed to disappear; since they are not deblended, the two doublets look like mini-BALs, and this is why we do not observe a disappearance;
- in the spectrum of J133152.19+051137.9 we find a trough whose measured width is below our threshold defining BALs ($\Delta v \geq 2000 \text{ km s}^{-1}$), and the difference between the threshold and the measured width is $< 1 \text{ km s}^{-1}$, nonetheless it cannot be considered as a BAL, technically;
- in the remaining three cases (J085904.59+042647.8, J094806.58+045811.7, J114546.22+032251.9) we do not detect any BAL troughs in the region

indicated in Filiz Ak et al. (2012) in the corresponding SDSS-I/II epoch because of the presence of a peak in the spectrum crossing the BAL-trough threshold (i.e., 90% of the normalized continuum level) upwards; as a consequence, what we see is two adjacent troughs, each having a width $\Delta v < 2000 \text{ km s}^{-1}$ but, if the peak were not there, we would get a BAL in each case.

In Figure 8.8 we show the pairs of spectra corresponding to each of the five mentioned QSOs.

In Filiz Ak et al. (2012) the fraction of disappearing BAL troughs is $f_{\text{disapp}_{F12}} = 21/925 = 2.3^{+0.6}_{-0.5}\%$, while the fraction of QSOs showing at least one disappearing BAL trough in their spectra is $f_{\text{QSO}_{F12}} = 19/582 = 3.3^{+0.9}_{-0.7}\%$; both percentages are a little lower than the ones we find for our P_4 sample, but are consistent with the percentages obtained from the analysis of our P_8 sample. The estimated average BAL-trough lifetime from Filiz Ak et al. (2012) is $\bar{t}_{\text{trough}_{F12}} = 109^{+31}_{-22}$, while the BAL-phase duration is $\bar{t}_{\text{BAL}_{F12}} = 150^{+60}_{-50}$; both are consistent with the estimates that we draw up both from the P_8 and the P_4 sample.

The velocity distributions for our P_8 sample and FA sample are in good agreement, as can be seen from Figures 8.9 and 8.10, where they are shown in order to allow the comparison. The probability values obtained from the K-S test performed by Filiz Ak et al. (2012) are listed in the caption. We note that the probability obtained for the Δv distribution is $P \approx 10^{-4}$ for the P_8 sample and $P \approx 10^{-1}$ for the FA sample; in the other cases, the discrepancy between the two values is much lower.

In Chapter 9 we gather all these results together to ease the comparison between the findings from the two works.

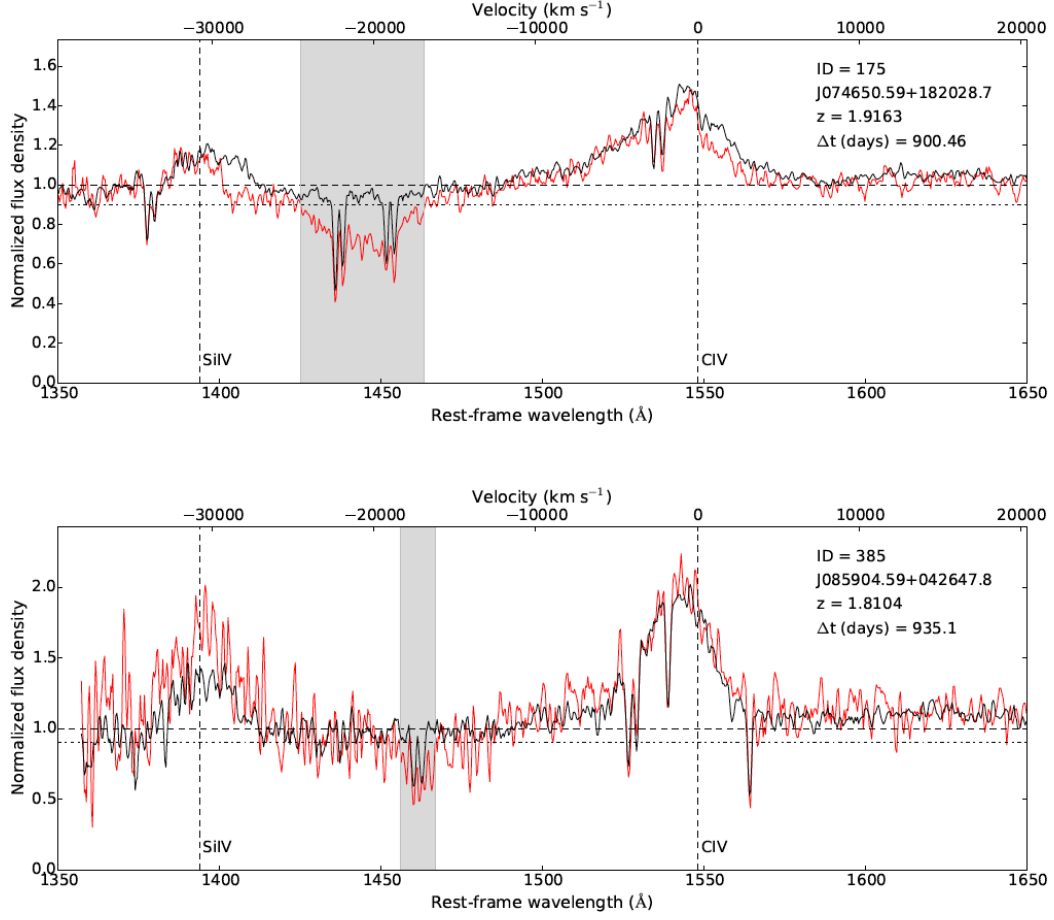


Figure 8.8: Overplot of the five SDSS-I/II (red) and BOSS (black) spectra where disappearing BAL troughs were detected in Filiz Ak et al. (2012), but not in the present work. On top right of each panel we report the ID of the corresponding source from our catalog, the SDSS ID, the redshift, and the rest-frame time difference between the two spectra represented. Plots are restricted to the range 1350 – 1650 Å for a better visualization of the window where C IV BAL disappearance can be observed. Rest-frame wavelength (bottom) and velocity (top) are reported on the horizontal axes, while the normalized flux density is shown on the vertical axis. The horizontal dashed line represents the level where normalized flux density equals unity, while the dash-dot line indicates the 0.9 level for the normalized flux density: our analysis concerns BAL troughs extending below this threshold. The two vertical dashed lines represent the rest-frame wavelengths corresponding to the Si IV and C IV emission lines (1394 Å and 1549 Å, respectively).

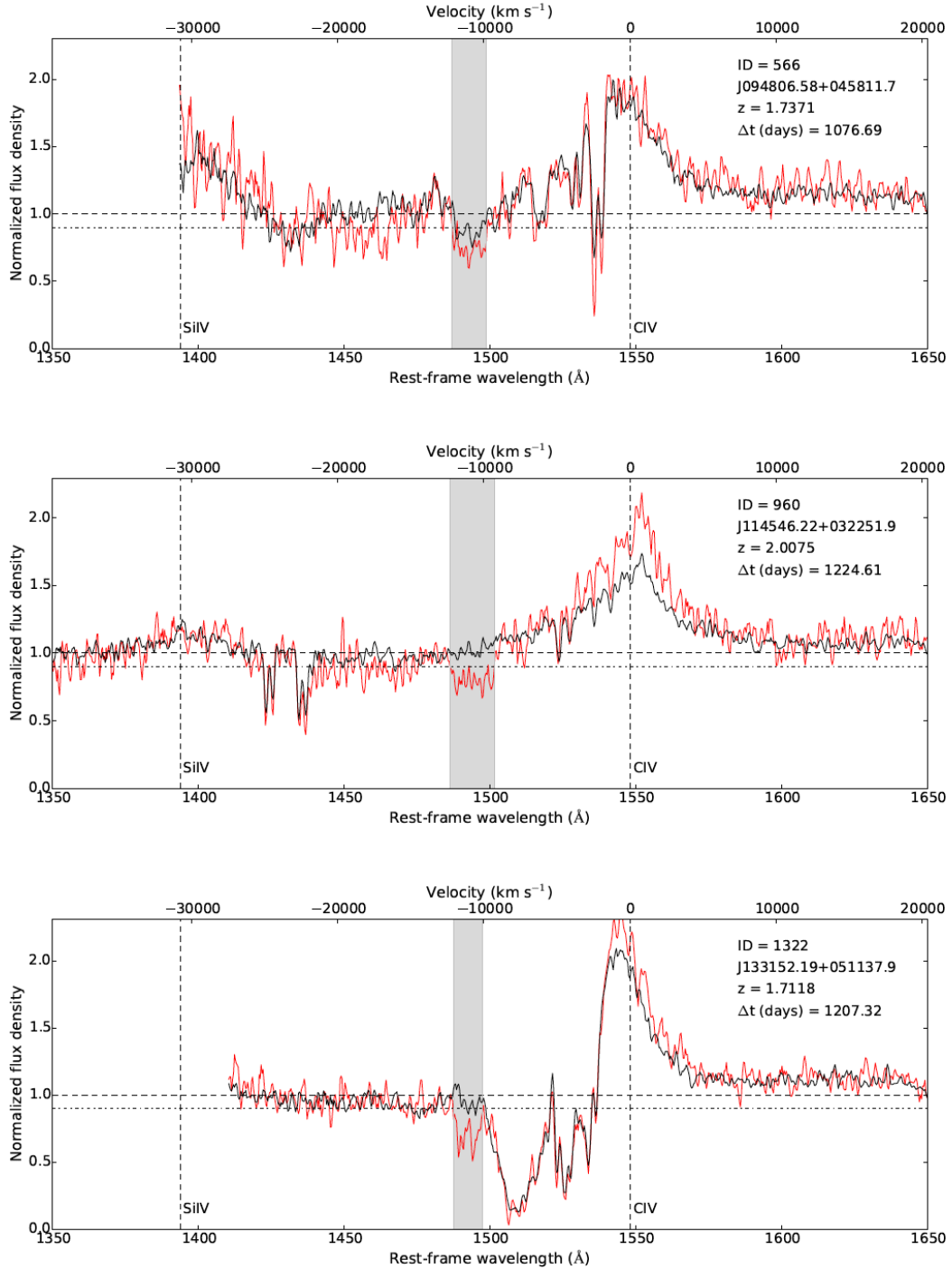


Figure 8.8: Overplot of the five SDSS-I/II (red) and BOSS (black) spectra where disappearing BAL troughs were detected in Filiz Ak et al. (2012), but not in the present work (cont.). We point out that, in the case of QSO J114546.22+032251.9, there are two peaks –and not just one (see Section 8.3)– in the SDSS-I/II spectrum crossing the threshold upwards, so we see three adjacent troughs making up the BAL; nevertheless, the first two corresponding to higher velocities are enough to make a BAL trough if we sum their widths.

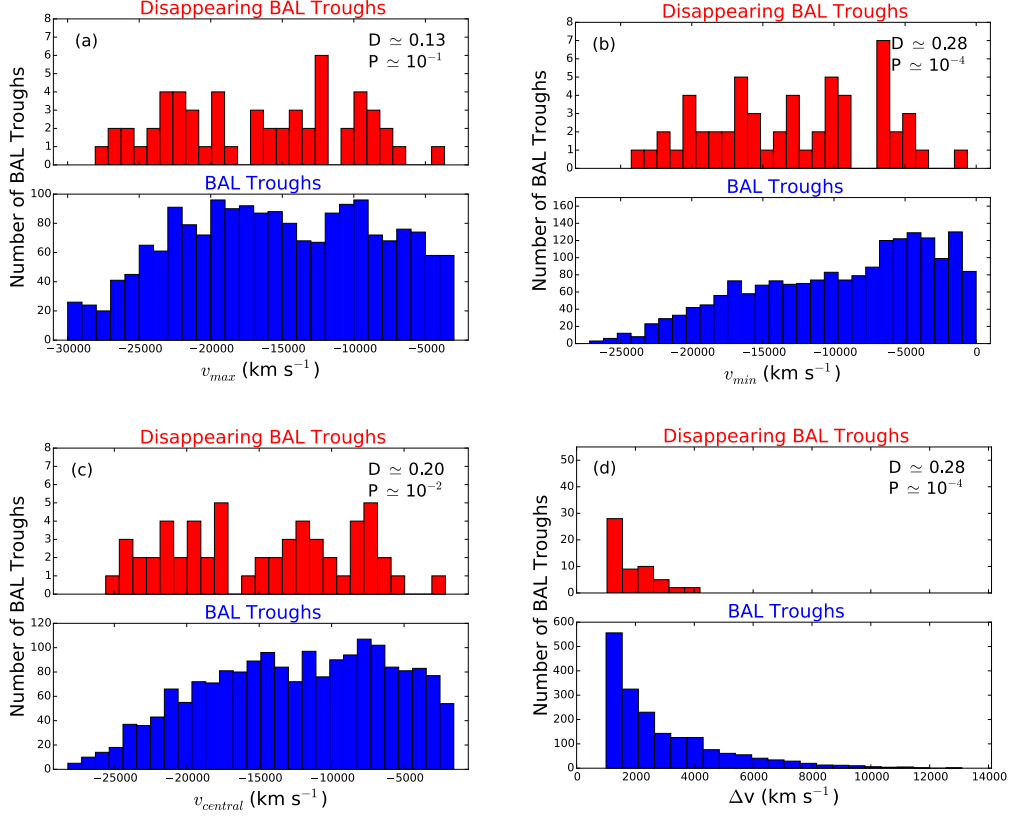


Figure 8.9: Maximum observed velocity v_{\max} (a), minimum observed velocity v_{\min} (b), central velocity v_c (c), and BAL width Δv (d) distributions for the P_8 sample (upper histogram in each panel) and for the main sample of sources with BAL troughs (lower histogram in each panel). Results of the K-S test performed on each pair of cumulative distributions are reported in each panel: D is the maximum distance between the two cumulative distributions, and P is the probability to get a higher D value assuming that the two datasets are drawn from the same distribution function. The corresponding distributions for the P_4 sample are shown in Figure 8.4, while in Figure 8.10 the distributions from Filiz Ak et al. (2012) are presented.

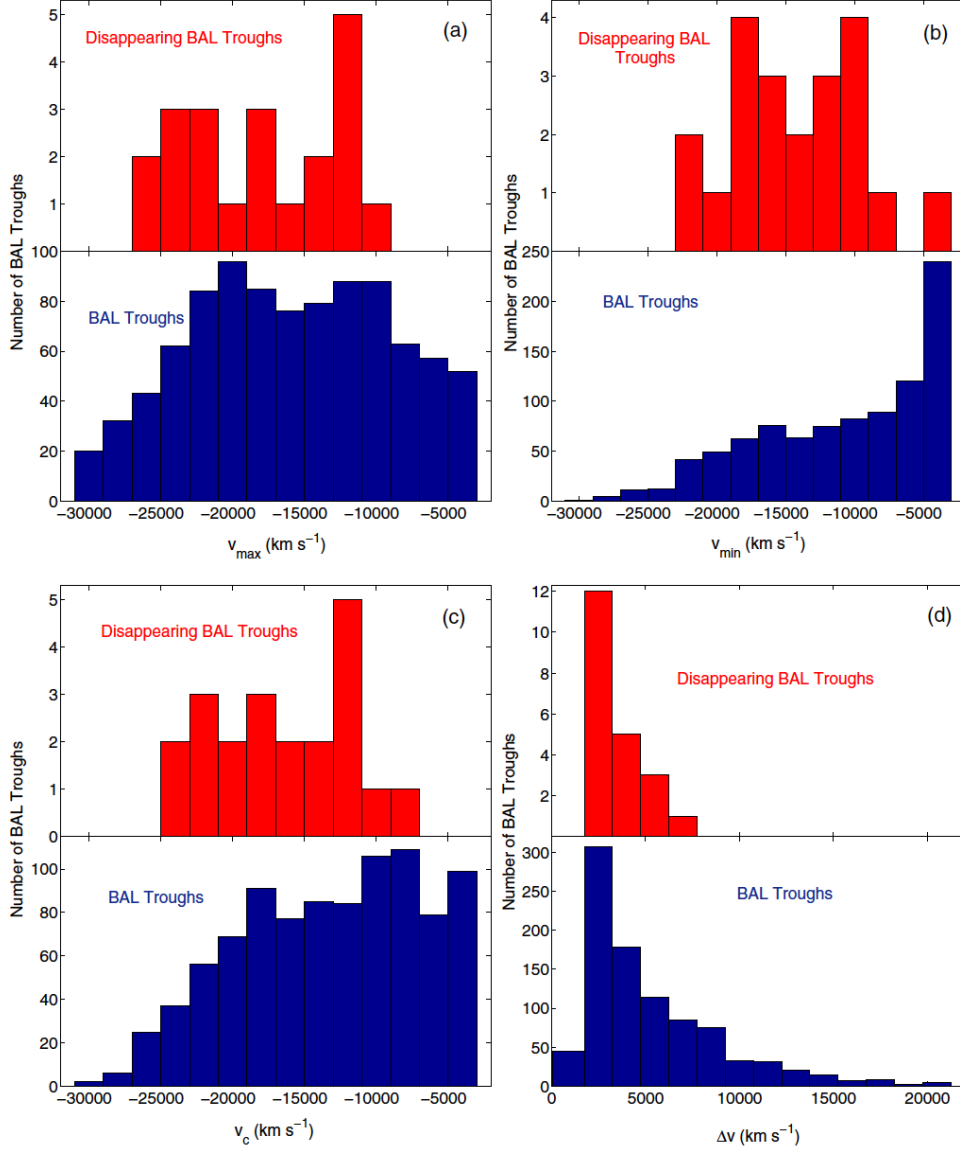


Figure 8.10: Maximum observed velocity v_{\max} (a), minimum observed velocity v_{\min} (b), central velocity v_c (c), and BAL width Δv (d) distributions for the FA sample (upper histogram in each panel) and the corresponding main sample (lower histogram in each panel), consisting of 582 sources where 925 C IV BAL troughs were detected (Filiz Ak et al., 2012). The probabilities obtained from the K-S test performed on each pair of cumulative distributions are: 10^{-1} (a), 10^{-3} (b), 10^{-2} (c), and 10^{-1} (d). The comparison with Figure 8.9, where the corresponding distributions for the P_8 sample are presented, shows that both distributions and probabilities are generally consistent, except for the two probabilities obtained for the Δv distributions.

Chapter 9

Discussion

In the second part of the present work we selected a sample of QSOs exhibiting C IV BAL troughs in their spectra and performed a statistical analysis of the subsample of BALs that disappear with the aim of extending our knowledge of the physical processes originating BAL variability and of the QSO structure as a whole.

The main sample of QSOs consists of 1319 sources and we detect 1874 C IV BAL troughs in their spectra; the sample of sources showing disappearing BALs consists of 67 QSOs, and the disappearing BALs are 73 (P_4 sample). Such disappearances are observed over a rest-frame timescale ranging 0.28 – 4.9 yr. The fraction of sources with a disappearing BAL is therefore $f_{\text{QSO } P_4} = 5.1^{+0.7}_{-0.6}\%$, while the fraction of disappearing BALs is $f_{\text{disapp } P_4} = 3.9^{+0.5}_{-0.5}\%$. On the basis of the frequency of disappearances, we can estimate the average BAL lifetime –limited to the direction of our line of sight– which turns out to be $\bar{t}_{\text{trough } P_4} \approx 80^{+10}_{-10}$ yr. Such an estimate is roughly consistent with the orbital time of the accretion disk at typical distances where winds are thought to form ($\approx 10^{-2}$ pc), the orbital time being on the order of 50 yr (e.g., Filiz Ak et al., 2013, and references therein): this suggests that a possible explanation

for BAL disappearance is disk rotation, as proposed in several works (e.g. Proga et al., 2000), and may mean that we observe disappearances because BALs move out of our line of sight while still existing physically.

Thirty of the BAL QSOs in our P_4 sample turn into non-BAL QSOs when BALs in their spectra disappear, hence we can compute the corresponding fraction of transforming BAL QSOs, which is $f_{\text{transform}_{P_4}} = 2.3^{+0.5}_{-0.4}\%$, and this allows us to get an estimate of the average lifetime of the BAL phase in a QSO, that is, $\bar{t}_{\text{BAL}_{P_4}} \approx 136^{+30}_{-24}$ yr; once again, the estimate is limited to what we can measure along our line of sight.

We also selected a more reliable sample of 56 disappearing BALs in the corresponding 52 sources, namely the P_8 sample, and the mentioned quantities for such a sample become, respectively: $f_{\text{QSO}_{P_8}} = 3.9^{+0.6}_{-0.5}\%$, $f_{\text{disapp}_{P_8}} = 3.0^{+0.5}_{-0.4}\%$, respectively, $\bar{t}_{\text{trough}_{P_8}} \approx 104^{+17}_{-14}$ yr, $f_{\text{transform}_{P_8}} = 1.8^{+0.5}_{-0.4}\%$, and $\bar{t}_{\text{BAL}_{P_8}} \approx 174^{+48}_{-39}$ yr.

We compared our findings with the results from Filiz Ak et al. (2012), where part of our sample of BAL QSOs was analyzed; all of them are reported in Table 9.1, to favor a straightforward comparison. It is apparent that all the results we obtain for our P_8 sample are consistent with the corresponding results from Filiz Ak et al. (2012); if we focus on the P_4 sample, this still holds, except for the first two fractions reported in the table, which are slightly larger in our analysis, even when we take into account error bars.

The distributions of minimum velocity, central velocity, and velocity difference for our main sample and for the P_4 sample (see Section 8.2.1), as well as the corresponding EW distributions (Section 8.2.2), show that the BAL troughs that disappear are generally narrow and characterized by a higher outflow velocity with respect to non-disappearing BALs.

The existence of a coordination in BAL variability is apparent from the

Table 9.1: Comparison between the main numerical findings from the present work and Filiz Ak et al. (2012).

	P_4 sample	P_8 sample	Filiz Ak et al. (2012)
Fraction of sources with disappearing BAL troughs	$5.1^{+0.7}_{-0.6}\%$ (67/1319)	$3.9^{+0.6}_{-0.5}\%$ (52/1319)	$3.3^{+0.9}_{-0.7}\%$ (19/582)
Fraction of disappearing BAL troughs	$3.9^{+0.5}_{-0.5}\%$ (73/1874)	$3.0^{+0.5}_{-0.4}\%$ (56/1874)	$2.3^{+0.6}_{-0.5}\%$ (21/925)
Average BAL-trough lifetime \bar{t}_{trough} (yr)	80^{+10}_{-10}	104^{+17}_{-14}	109^{+31}_{-22}
Fraction of BAL QSOs that turn into non-BAL QSOs	$2.3^{+0.5}_{-0.4}\%$ (30/1319)	$1.8^{+0.5}_{-0.4}\%$ (24/1319)	$1.7^{+0.7}_{-0.5}\%$ (10/582)
Average BAL-phase lifetime \bar{t}_{BAL} (yr)	136^{+30}_{-24}	174^{+48}_{-39}	150^{+60}_{-50}

plots shown in Section 8.2.2: in spectra where more than one BAL is detected and one of them disappears, the others get weaker in 79% of the cases, while in the main sample population there is no dominant trend between strengthening and weakening. Also, in 96% of the cases the disappearing BAL is the one with the highest outflow velocity. Coordination in variability is observable even for BAL troughs that are very distant from each other (central velocity offset up to $\approx 20000 \text{ km s}^{-1}$).

In Section 1.5 we mentioned that the observed BAL trough variability could be caused by variations in the covering factor which originate from the motion of the gaseous clouds along our line of sight; nevertheless, the cause of variability coordination in multiple BALs at different velocities in the same spectrum must be other than this, as BALs arising at different velocities correspond to different radial distances from the central SMBH and therefore originate in gaseous regions that are separated from one another (e.g., Capellupo et al., 2012).

The explanation must hence be global, rather than local; a plausible one attributes coordinated variability to changes in the ionization level of the absorbing gas, originating from changes in the ionizing flux impacting on the gas itself, which could be in turn ascribed to variations in the column density of the shielding gas. Such changes affect the outflow as a whole, thus giving rise to coordinated variations in the absorption troughs. Anyway, we should keep in mind that more saturated lines are scarcely responsive to changes in the ionization level, while changes in the covering factor can play a role in BAL variations; it is therefore likely that both causes contribute to the observed phenomenon, and the combined effect could be an enhanced variability or, in some other cases, a partial balance (e.g., Capellupo et al., 2012; Filiz Ak et al., 2012).

Conclusions

The present thesis investigated AGN variability from two distinct perspectives. In the first part we illustrated the strength of a method based on optical variability to select AGN candidate samples, applied to data from the VST surveys of the COSMOS field and the CDFS. We pushed towards deeper magnitude values than in past studies, and compared our results to the ones obtained by means of different selection techniques to identify AGNs, discussed in other recent works from the literature. The selection of AGN candidates on the basis of their optical variability allowed to obtain a high-purity (83%) sample, which was confirmed by means of a number of multi-wavelength diagnostics. Moreover, the extension of the analysis from a five month to a three year baseline led to a rise in the completeness with respect to the X-ray confirmed AGNs (from 15% to 41%). This demonstrated the reliability and effectiveness of the method and proved it a powerful tool to be used when, in the near future, the reduction and analysis of much larger amounts of data coming from wide-field surveys (such as LSST) will be required.

We also investigated the trend of AGN variability in different BH mass ranges and accretion rate ranges by computing the structure function of our X-ray confirmed AGNs: our results are still preliminary and need some refinement; nonetheless, we can state they support (consistently with other works

from the literature) the existence of an anti-correlation with the accretion rate, while no relation with the BH mass emerges.

The availability of additional VST data for the CDFS in the u , g , r , and i bands will allow us to make use of multi-band variability to refine our AGN selection criteria and obtain reliable estimates of the physical parameters characterizing the emitting region.

The second part of the thesis was oriented towards a deeper understanding of the physics and structure of QSOs in the more general context of their formation and evolution; this was pursued through the analysis of the disappearance of C IV BALs in the largest sample of BAL QSOs ever investigated so far, coming from SDSS-I/II/III surveys. The analysis of C IV BAL disappearance quantified the frequency of the phenomenon and provided estimates of the lifetime of BAL troughs, which turned out to be roughly consistent with the orbital time of the accretion disk at typical distances from the center where winds are thought to originate, thus supporting the idea that BAL disappearance may be caused by disk rotation.

The analysis also provided evidence for the existence of a coordination in the variability of multiple troughs corresponding to the same transition, which persists even when the radial distances between the two BALs are very large (corresponding to velocity offsets on the order of 10^4 km s^{-1}). BALs are thought to form because of outflowing winds originating in the proximity of the central SMBH, and the existence of a coordination in their variability helps shed light onto the possible mechanisms behind BAL formation and variability itself, suggesting variations are plausibly caused by a change in the features (such as density) of the shielding gas between the source of radiation and the region where winds arise. It is likely that both rotation and variations in the shielding gas are responsible for the BAL variability

phenomenon as a whole.

We are planning to extend the analysis to lower ionization transitions – e.g., Si IV– in future works, as this would allow us to study additional samples of BAL troughs and to investigate possible relations between the variability of troughs corresponding to different transitions; also, since new spectra for our sample of BAL QSOs are currently being obtained by the Time Domain Spectroscopic Survey (TDSS; e.g., Morganson et al. 2015), the analysis of re-emergence of previously disappeared BALs could be performed for those sources for which at least three epochs are available: this would be a further, significant step towards a deeper knowledge of the BAL phenomenon and could help us put additional –and possibly tighter– constraints on the physics of BAL formation, evolution, and variability.

Bibliography

- Abbott, T., & the DES collaboration, *MNRAS*, 2016, **460**, 1270.
- Abell, P. A., & the LSST collaboration, *LSST Science Book*, 2009, <https://arxiv.org/pdf/0912.0201.pdf>.
- Aihara, H., et al., *ApJS*, 2011, **193**, 29.
- Andersen, M. I., et al., in *Calibrating and Understanding HST and ESO Instruments*, ed. P. Benvenuti, *European Southern Observatory Conference and Workshop Proc.*, 1995, **53**, 87.
- Antonucci, R., *ARAA*, 1993, **31**, 473.
- Antonucci, R. R. J. & Miller, J. S., *ApJ*, 1985, **297**, 621-632.
- Arexaga, I., & Terlevich, R., *MNRAS*, 1994, **269**, 462.
- Baldwin, J. A., Phillips, M. M., & Terlevich, R., *PASP*, 1981, **93**, 5.
- Bamford, S. P., et al, *MNRAS*, 2008, **391**, 607.
- Barlow, T. A., 1993, PhD Thesis, California University.
- Barr, P., & Mushotsky, R. F., *Nature*, 1986, **320**, 421.
- Bauer, A., et al, *ApJ*, 2009, **696**, 1241.

- Becker, R. H., et al, *ApJ*, 1997, **479**, L93.
- Becker, R. H., et al, *ASPC*, 1997, **128**, 31.
- Bershady, M. A., et al, *ApJ*, 1998, **496**, 103.
- Bertin, E., & Arnouts, S., *A&AS*, 1996, **117**, 393.
- Bertin, E., et al., in *Astronomical Data Analysis Software and Systems XI*, eds. D. A. Bohlender, D. Durand, & T. H. Handley, *ASP Conf. Ser.*, 2002, **281**, 228.
- Bianchi, S., et al. *AdAst*, 2012, 17.
- Botticella, M. T., et al., *A&A*, 2017, **598**, A50.
- Botticella, M. T., et al., *The Messenger*, 2013, **151**, 29.
- Brandt, W. N., & Hasinger, G., *ARA&A*, 2005, **43**, 827.
- Brotherton, M. S., et al., *AJ*, 2005, **130**, 2006.
- Brotherton, M. S., et al., *ApJ*, 1998, **50**, L7.
- Brusa, M., et al., *ApJ*, 2010, **716**, 348.
- Capaccioli, M., & Schipani, P., *The Messenger*, 2011, **146**, 2.
- Capaccioli, M., et al., *A&A*, 2015, **581**, A10.
- Capak, P., et al., *ApJS*, 2007, **172**, 99.
- Capellupo, D. M., et al., *MNRAS*, 2012, **422**, 3249.
- Capetti, A., & Baldi, R. D., *A&A*, 2011, **529**, A126.
- Cappellaro, E., et al., *A&A*, 2015, **584**, A62.

- Cardelli, J. A., et al., *ApJ*, 1989, **345**, 245.
- Civano, F., et al., *ApJ*, 2016, **819**, 62.
- Cristiani, S., et al., *A&A*, 1997, **321**, 123.
- Cristiani, S., et al., *A&A*, 1996, **306**, 395.
- Dawson, K. S., et al., *AJ*, 2013, **145**, 10.
- De Cicco, D., et al., *A&A*, 2015, **574**, A112.
- De Cicco, D., et al., in preparation.
- de Vries, W. H., et al., *AJ*, 2005, **129**, 615.
- Di Clemente, A., et al., *ApJ*, 1996, **463**, 466.
- Di Matteo, T., et al., *Nature*, 2005, **433**, 604.
- Donley, J. L., et al., *ApJ*, 2012, **748**, 142.
- Eisenstein, D. J., et al., *AJ*, 2011, **142**, 72.
- Elvis, M., *ApJ*, 2000, **545**, 63.
- Elvis, M., *ApJS*, 1994, **95**, 1.
- Falocco, F., et al., *A&A*, 2015, **579**, A115.
- Fan, X., *AJ*, 1999, **117**, 252.
- Fazio, G. G., et al., *ApJS*, 2004, **154**, 10.
- Ferrarese, L., & Merritt, D., *ApJ*, 2000, **539**, L9.
- Filiz Ak, N., et al., *ApJ*, 2013, **777**, 168.

- Filiz Ak, N., et al., *ApJ*, 2012, **757**, 114.
- Fiore, F., et al., *ApJ*, 2009, **693**, 447.
- Fossati, G., et al., *MNRAS*, 1998, **299**, 433.
- Gaskell, C. M. & Klimek, E. S., *A&AT*, 2003, **22** (4-5), 661.
- Gehrels, N., *ApJ*, 1986, **303**, 336.
- Gibson, R. R., et al., *ApJ*, 2009, **692**, 758.
- Gibson, R. R., et al., *ApJ*, 2008, **675**, 985.
- Giveon, U., et al., *MNRAS*, 1999, **306**, 637.
- Grado, A., et al., *MSAIS*, 2012, **19**, 362.
- Graham, M. J., et al., *MNRAS*, 2014, **439**, 703.
- Green, P. J., et al., *ApJ*, 2001, **558**, 109.
- Grier, C. J., et al., *ApJ*, 2016, **824**, 130.
- Grier, C. J., et al., *ApJ*, 2015, **806**, 111.
- Gültekin, K., et al., *ApJ*, 2009, **698**, 198.
- Gunn, J. E., et al., *AJ*, 2006, **131**, 2332.
- Hall, P. B., et al., *MNRAS*, 2013, **434**, 222.
- Hall, P. B., et al., *ApJS*, 2002, **141**, 267.
- Hazard, C., et al., *ApJ*, 1987, **323**, 263.
- Heckman, T. M., *A&A*, 1980, **87**, 152.

- Hewett, P. C., & Wild, V., *MNRAS*, 2010, **405**, 2302.
- Hsu, L.-T., et al., *ApJ*, 2014, **796**, 60.
- Huang, Z., et al., *A&A*, 2011, **529**, A93.
- Hughes, P. A., et al., *ApJ*, 1992, **396**, 469.
- Ilbert, O., et al., *ASPC*, 2008, **399**, 169.
- Klesman, A., & Sarajedini, V., *ApJ*, 2007, **665**, 225.
- Koekemoer, A. M., et al., *ApJS*, 2007, **172**, 196.
- Kormendy, J., & Ho, L. C., *ARA&A*, 2013, **51**, 511.
- Kormendy, J., & Richstone, D., *ARA&A*, 1995, **33**, 581.
- Krolik, J. H., *Active Galactic Nuclei. From the Central Black Hole to the Galactic Environment*, Princeton University Press, 1999.
- Kuijken, K., *The Messenger*, 2011, **146**, 8.
- Lacy, M., et al., *AJ*, 2007, **133**, 186.
- Lacy, M., et al., *ApJS*, 2004, **154**, 166.
- Landt, H., et al., *ApJS*, 2008, **174**, 282.
- Lawrence, A., & Papadakis, I., *ApJ*, 1993, **414**, L85.
- Lehmer, B. D., et al., *ApJS*, 2005, **161**, 21.
- Lilly, S. J., et al., *ApJS*, 2007, **172**, 70.
- Lonsdale, C., et al., *ApJS*, 2004, **154**, 54.

- Lundgren, B. F., et al., *ApJ*, 2007, **656**, 73.
- Lusso, E., et al., *MNRAS*, 2012, **425**, 623.
- Lusso, E., et al., *A&A*, 2010, **512**, A34.
- Maccacaro, T., et al., *ApJ*, 1988, **326**, 680.
- Magorrian, J., et al., *AJ*, 1998, **115**, 2285.
- Mainieri, V., et al., *A&A*, 2002, **393**, 425.
- Marchesi, S., et al., *ApJ*, 2016, **817**, 34.
- Margala, D., et al., *ApJ*, 2016, **831**, 157.
- Mauduit, J.-C., et al., *PASP*, 2012, **124**, 1135.
- McCracken, H. J., et al., *ApJ*, 2010, **708**, 202.
- McHardy, I. M., et al., *Nature*, 2006, **444**, 730.
- Morganson, E., et al., *ApJ*, 2015, **806**, 244.
- Murray, N., et al., *ApJ*, 1995, **451**, 498.
- Mushotzky, R., *ASSL*, 2004, **308**, 53.
- Nakos, Th., et al., *A&A*, 2009, **494**, 579.
- Netzer, H., *ARA&A*, 2015, **53**, 365.
- Paolillo, M., et al., *ApJ*, 2004, **611**, 93.
- Paolillo, M., et al., in preparation.
- Pei, Y. C., *ApJ*, 1992, **395**, 130.

- Pereyra, N. A., et al., *ApJ*, 2006, **642**, 87.
- Peters, C. M., et al., *ApJ*, 2015, **811**, 95.
- Peterson, B. M., *An Introduction to Active Galactic Nuclei*, Cambridge University Press, 1997.
- Peterson, B. M., et al., *ApJ*, 1998, **501**, 8.
- Poulain, M., et al., in preparation.
- Proga, D., et al., *ApJ*, 2000, **543**, 686.
- Rangel, C., et al., *MNRAS*, 2013, **428**, 3089.
- Richards, G. T., et al., *ApJS*, 2006, **16**, 470.
- Richards, G. T., et al., *AJ*, 2001, **121**, 2308.
- Rochais, T. B., et al., *MNRAS*, 2014, **444**, 2498.
- Rowan-Robinson, M., et al., *AJ*, 2005, **129**, 1183.
- Salvato, M., et al., *ApJ*, 2011, **742**, 61.
- Sanders, D. B., et al., *ApJS*, 2007, **172**, 86.
- Sbarrato, T., et al., *MNRAS*, 2014, **445**, 81.
- Schlegel, D. J., et al., *ApJ*, 1998, **500**, 525.
- Schmidt, K. B., et al., *ApJ*, 2012, **744**, 147.
- Schmidt, K. B., et al., *ApJ*, 2010, **714**, 1194.
- Schneider, D. P., et al., *AJ*, 2007, **134**, 102.

- Scoville, N., et al., *ApJS*, 2007, **172**, 38.
- Simm, T., et al., *A&A*, 2016, **585**, A129.
- Simonetti, J. H., et al., *ApJ*, 1985, **296**, 46.
- Stasinska, G., et al., *MNRAS*, 2008, **391**, L29.
- Sutherland, W., et al., *A&A*, 2015, **575**, A25.
- Trevese, D., et al., *A&A*, 2008, **488**, 73.
- Trouille, L., Barger, A. J., & Tremonti, C., *ApJ*, 2011, **742**, 4.
- Trump, J. R., et al., *ApJS*, 2006, **165**, 1.
- Ulrich, M.-H., et al., *ARA&A*, 1997, **35**, 445.
- Urry, C. M., & Padovani, P., *PASP*, 1995, **107**, 803.
- Uttley, P., *ASP Conf. Ser.*, 2006, **360**.
- Vaccari, M., et al., in preparation.
- Vanden Berk, D. E., et al., *ApJ*, 2004, **601**, 692.
- Villfort, C., et al., *ApJ*, 2010, **723**, 737.
- Virani, S. N., et al., *AJ*, 2006, **131**, 2373.
- Voit, G. M., et al., *ApJ*, 1993, **413**, 95.
- Webb, W., & Malkan, M., *ApJ*, 2000, **540**, 652.
- Weymann, R. J., et al., *ApJ*, 1991, **373**, 23.
- Wildy, C., et al., *MNRAS*, 2014, **437**, 1976.

Wood, S. N., et al., *Roy. Stat. Soc. B*, 2011, **73**, 3.

Xue, Y. Q., et al., *ApJS*, 2011, **195**, 10.

Yang, G., et al., *ApJ*, 2016, **831**, 145.

York, D. G., et al., *AJ*, 2000, **120**, 1579.

Young, M., et al., *ApJ*, 2012, **748**, 124.